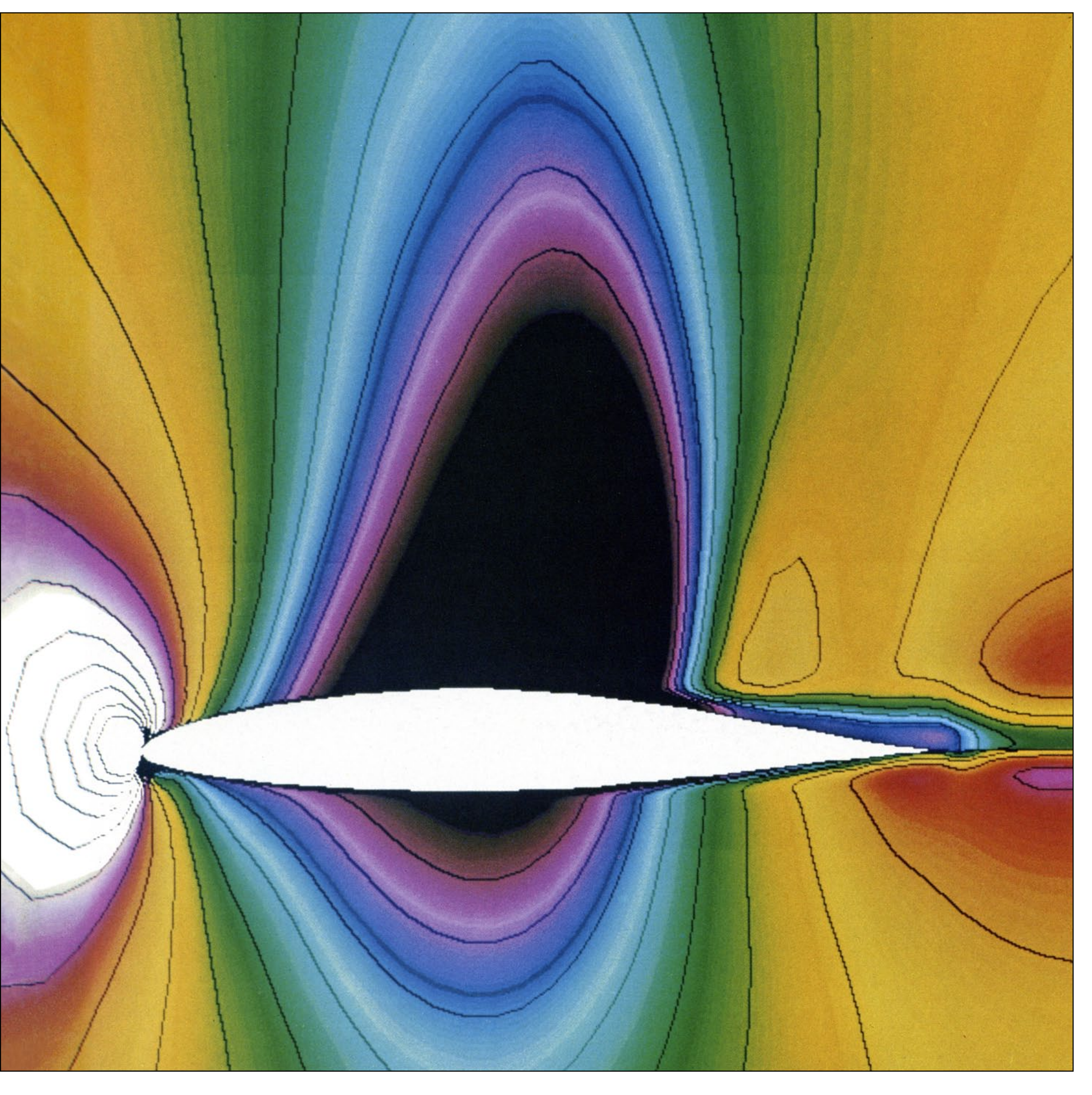
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



SUPERCOMPUTADORES

Marzo 1982

300 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 66
Marzo 1982

INVESTIGACION Y CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

EDITA: Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

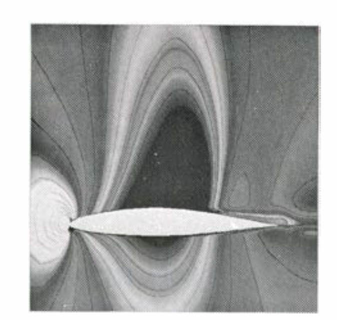
Mireia Bofill Abelló: *El sistema estadounidense de Seguridad Social*; Jaume Gallemí: *Extinciones en masa del Mesozoico tardío*; Pedro Pascual de Sans: *Estabilización del hidrógeno atómico*; Manuel Puigcerver: *Los satélites de Saturno*; F. J. López Aligué: *Supercomputadores*; Cristóbal Mezquita: *Formación de un sistema nervioso simple*; José María Tous y Juan Delgado: *Psicología de las preferencias*; Luis Bou: *Temas metamágicos*; J. Vilardell: *Taller y laboratorio*.

Ciencia y sociedad:

Julio Samsó, A. Palanca Soler, C. Castán Lanaspa y Sebastián Xambó

Libros:

Ramón Margalef, Emilio Lledó, Luis Alonso y Antoni M. Correig



LA PORTADA

La ilustración de la portada es un fotograma de una película que muestra el flujo sobre la superficie de un ala a 0,7 Mach, es decir, a 0,7 veces la velocidad del sonido, simulado con el supercomputador ILLIAC IV (véase "Supercomputadores", por Ronald D. Levine, en este mismo número). Las vibraciones de los alerones, objeto del estudio, son un efecto destructivo que puede presentarse al interaccionar las ondas de choque que se propagan por las alas con el movimiento de los alerones. El estudio numérico puso de relieve las variaciones de la densidad del aire para una red bidimensional de cerca de 7000 puntos distribuidos sobre la superficie del ala. Cada fotograma de la película precisó 40 millones de operaciones aritméticas, equivalentes a 2 segundos de tiempo de ejecución del ILLIAC IV. Las curvas coloreadas representan los contornos de densidad constante del aire. En amarillo oscuro aparece la densidad del aire no perturbado; los colores más claros indican zonas de compresión, y, los más oscuros, de depresión. La onda de choque provoca la aparición de una región de circulación inversa (*púrpura*). El cáculo lo realizaron Harry E. Bailey, Tony Hasegawa y Alan Johnson.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 37

Condiciones de suscripción:
España:
Un año (12 números): 3.300 pesetas
Extranjero:
Un año (12 números): 52 U.S. $
Ejemplar atrasado ordinario:
340 pesetas
Ejemplar atrasado extraordinario:
440 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49, Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Lourdes Padrós
Manuel Girona, 61, 2.°, Barcelona-34
Teléfono 204 45 83

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

La ilustración de portada es cortesía del Ames Research Center y la National Aeronautics and Space Administration

Página	Fuente
7-13	Ilil Arbel
17-24	Mercedes Alcover Ibañez Luis Vila López y Miguel Alonso
26	Walter Alvarez
28	Helen V. Michel (*arriba*) D. M. Jarzen (*abajo*)
29-33	Patricia J. Wynne
37	Julio Samsó
39	A. Palanca y C. Castán
43-51	Gabor Kiss
52	Jet Propulsion Laboratory, Instituto de Tecnología de California
53-55	Walken Graphics
56-60	Jet Propulsion Laboratory
62-63	Servicio Geológico de los Estados Unidos, Flagstaff, Arizona
66	Jet Propulsion Laboratory
67	Walken Graphics
71	Ames Research Center, National Aeronautics and Space Administration
72-73	Jon Brenneis
74-82	Allen Beechel
84	Ames Research Center, National Aeronautics and Space Administration
87	Saul L. Zackson, Universidad de California en Berkeley
88-89	Tom Prentiss
90-91	Roy T. Sawyer, Universidad de California
92	Andrew P. Kramer, Universidad de California en Berkeley (*izquierda*); Duncan K. Stuart, Universidad de California en Berkeley (*derecha*)
93	Andrew P. Kramer, Universidad de California
94	Saul L. Zackson, Universidad de California en Berkeley
95	Gunther S. Stent y David A. Weisblat
96	Tom Prentiss
101-105	Ilil Arbel
115	Marshall Harrington Photography
116-120	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Pedro IV, 160 - Barcelona-5
Fotocromos reproducidos por GINSA, S.A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Nápoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

ELI GINZBERG ("El sistema estadounidense de Seguridad Social") dirige el Proyecto de Conservación de los Recursos Humanos de la Universidad de Columbia. Su trato con este centro superior comenzó hace más de medio siglo. Allí se matriculó, licenció luego y se doctoró, por último, en 1934. Se incorporó al claustro de profesores en 1935 y ocupó la cátedra A. Barton Hepburn de Economía en la Graduate School of Business de Columbia desde 1967 hasta 1979, año en que se jubiló; sigue dictando ahora cursos especiales. Además de atender a sus responsabilidades académicas, ha intervenido en numerosas ocasiones como asesor de varios altos funcionarios y organismos gubernamentales. Ginzberg desea hacer constar aquí la ayuda recibida de sus colegas Morton Miller, Daniel Saks, Steven Sandell y Anna B. Dutka en la preparación de este artículo.

MERCEDES ALCOVER IBAÑEZ y LUIS VILA LOPEZ ("El sistema español de Seguridad Social") han trabajado en encuestas de opinión sobre temas de Sanidad y Seguridad Social. Autores, conjuntamente, de artículos diversos sobre sociología de la salud y del trabajo, forman parte de un servicio de estudios que ha trabajado para la Administración. Alcover Ibáñez es licenciada en ciencias políticas (Madrid, 1970) con estudios de postgraduado en Madison (Wisconsin) y especialización en relaciones industriales y organización del trabajo. Vila López es economista (Barcelona, 1974); ha sido profesor de sociología en las Universidades de Barcelona y Alcalá de Henares. Investiga sobre políticas de bienestar y calidad de vida.

DALE A. RUSSELL ("Extinciones en masa del Mesozoico tardío") es conservador de vertebrados fósiles y jefe de la división de paleontología del Museo Nacional de Ciencias Naturales del Canadá, en Ottawa. Licenciado por la Universidad de Oregón, cursó el segundo ciclo en la de California en Berkeley y se doctoró por la de Columbia. Entre sus líneas de investigación se cuentan "la ecología de faunas reptiles terrestres a finales de la edad de los reptiles, la evolución del cerebro en los dinosaurios y tendencias en la evolución de organismos pluricelulares que puedan ser importantes en el desarrollo de ecosistemas extraterrestres".

ISAAC F. SILVERA y JOOK WALRAVEN ("Estabilización del hidrógeno atómico") son físicos experimentales de la Universidad de Amsterdam. Silvera nació y se educó en los Estados Unidos, recibiendo su doctorado en física por la Universidad de California en Berkeley, en 1965. Después de pasar un año en la Universidad de Grenoble, trabajó para el Rockwell International Science Center de California hasta 1971, cuando fue nombrado profesor de física experimental en la Universidad de Amsterdam. La investigación llevada a cabo por Silvera se centra principalmente en la aplicación de técnicas espectroscópicas al estudio de sistemas mecánico-cuánticos a bajas temperaturas y altas presiones. Walrawen se ha dedicado al estudio del hidrógeno atómico en el estado condensado desde 1972, primero durante su licenciatura y después como alumno de doctorado subvencionado por la Fundación para la Investigación Fundamental de la Materia, de Utrecht.

LAURENCE A. SODERBLOM y TORRENCE V. JOHNSON ("Los satélites de Saturno") son especialistas en ciencias planetarias con un interés particular en los satélites del sistema solar. Soderblom es jefe de la rama de estudios astrogeológicos de la Inspección Geológica de los Estados Unidos. Se licenció en geología y física por el Instituto de Minería y Tecnología de Nuevo México y se recibió de doctor, en ciencias planetarias y geofísica, por el Instituto de Tecnología de California. Desde su ingreso en la Inspección Geológica, en 1970, ha participado en varias misiones no tripuladas de exploración planetaria lanzadas por la Administración Nacional de Aeronáutica y del Espacio (NASA), incluyendo las misiones Voyager a los planetas exteriores. Soderblom es el autor de "Los satélites galileanos de Júpiter", que apareció en el número de marzo de 1980 de INVESTIGACIÓN Y CIENCIA. Johnson explica ciencias planetarias en el Instituto de Tecnología de California y trabaja como investigador científico en el Laboratorio de Propulsión a Chorro.

RONALD D. LEVINE ("Supercomputadores") trabaja en la compañía Technology Development of California, Inc., donde se ocupa de varios aspectos del cálculo de problemas científicos bajo contrato de la NASA. Cursó los dos ciclos universitarios en la de California en Berkeley, doctorándose en matemáticas por esa institución en 1970. Ha estudiado, también, en las universidades de Stanford, Cornell y Cambridge. Durante la década de 1960 trabajó en el Lawrence Berkeley Laboratory, en el Stanford Linear Accelerator Center y en el Jet Propulsion Laboratory del Instituto de Tecnología de California. Cuenta de sí mismo: "Durante los últimos 20 años me he interesado por muchos aspectos científicos, desde la teoría de la relatividad hasta la biología matemática, con los problemas de cómputo como nexo de unión entre tan diversos campos".

GUNTHER S. STENT y DAVID A. WEISBLAT ("Formación de un sistema nervioso simple") trabajan en la Universidad de California en Berkeley, donde Stent dirige el departamento de biología molecular y laboratorio de virus, y Weisblat completa su preparación postdoctoral. Stent nació en Berlín y se trasladó a los Estados Unidos en 1940. Se licenció y doctoró en fisicoquímica por la Universidad de Illinois y realizó su trabajo postdoctoral ("en la disciplina que con el tiempo sería conocida como biología molecular") en el Instituto de Tecnología de California (Cal Tech) y en el Instituto Pasteur. Se trasladó a Berkeley en 1952. Sus investigaciones sobre la neurobiología de la sanguijuela se iniciaron en 1969, durante un año sabático de estancia en la Facultad de Medicina de Harvard, trabajando con Stephen W. Kuffler y John G. Nicholls. Weisblat se licenció por Harvard y se doctoró en química y neurofisiología por el Cal Tech.

DANIEL KAHNEMAN y AMOS TVERSKY ("Psicología de las preferencias") comenzaron su colaboración en 1969, en la Universidad Hebrea de Jerusalén. Abandonaron Israel en 1978 para ocupar cargos en instituciones académicas de Norteamérica, Kahneman como profesor de psicología en la Universidad de British Columbia y Tversky como profesor de psicología en la de Stanford. Kahneman, que nació en Israel, se licenció y doctoró en psicología por la Universidad de California en Berkeley en 1961. Trabajó en las de Michigan y Harvard, en el Instituto de Investigación de Oregón, en la Universidad de Investigación en Psicología Aplicada de Cambridge, y en el Centro para el Estudio Avanzado en Ciencias de la Conducta de Stanford. Tversky se doctoró por la Universidad de Michigan en 1965.

Hace...

José M.ª López Piñero

... cien años

José Gómez Ocaña obtuvo el título de licenciado en la Facultad de Medicina de Granada. Nacido en Málaga el año 1860, durante sus estudios universitarios tuvo entre sus maestros a tres excelentes cultivadores de las ciencias morfológicas aplicadas a la medicina. El primero de ellos fue Aureliano Maestre de San Juan, introductor en nuestro país de la moderna histología. El segundo, Juan Creus Manso, el más destacado tratadista español de anatomía topográfica. El tercero, Benito Hernando, en cuyos trabajos leprológicos participaron Rudolf Virchow, André Victor Cornil y Albert Neisser, como hace poco expusimos en estas mismas páginas. Por el contrario, la formación de Gómez Ocaña en fisiología, disciplina a la que se dedicó su vida, partió del más completo autodidactismo, ya que la correspondiente cátedra granadina ni siquiera contaba entonces con titular.

Una vez graduado, intentó ejercer como médico rural, pero su nula inclinación a la práctica clínica le hizo fracasar. En 1885 se trasladó a Madrid, en cuya Facultad de Medicina había ganado una plaza de ayudante. Durante un año vivió consagrado a sus deberes universitarios, a dar clases particulares y a la preparación de oposiciones. A mediados de 1886 ganó las correspondientes a la cátedra de fisiología de la Facultad de Medicina de Cádiz, que ocupó ocho años. Aunque los medios que allí tenía eran extraordinariamente modestos, gracias a su dedicación exclusiva a la disciplina pudo enseñarla de forma práctica en el laboratorio y mantener a sus discípulos al día de las últimas novedades europeas. Publicó entonces sus primeras monografías, tituladas *Fisiología del cerebro* (1894) y *Fisiología de la circulación en el organismo humano con aplicaciones a la patología y a la terapéutica* (1894), una serie de artículos de síntesis sobre fisiología digestiva en la *Gaceta Médica de Cádiz* (1893-94), así como algunos trabajos de investigación.

En 1894 se trasladó a la cátedra de fisiología de Madrid, de la que fue titular hasta su muerte. Con un laboratorio algo mejor dotado, continuó en la misma línea que en Cádiz. Aparte de realizar una fecunda labor docente, publicó una monografía sobre el tiroides (1895), un tratado de fisiología (1896) –que alcanzó cinco ediciones en dos décadas– y un elevado .número de trabajos experimentales. Algunos de ellos aparecieron en revistas médicas españolas, principalmente *El Siglo Médico* y la *Revista Iberoamericana de Ciencias Médicas*, que dirigía Federico Rubio. Otros lo hicieron en los anales y boletines de varias instituciones científicas de las que fue miembro, sobre todo las Reales Academias de Medicina y de Ciencias y otras de carácter más renovador, como la Asociación Española para el Progreso de las Ciencias y la Sociedad Española de Biología.

Especial interés tuvo Gómez Ocaña por presentar el resultado de sus trabajos experimentales en los congresos internacionales de fisiología y de medicina. Participó en ellos de modo regular desde su época de Cádiz, perteneciendo incluso al comité organizador de los congresos internacionales de fisiólogos que se celebraron en Heidelberg (1907), Viena (1910) y Groninga (1913). Sus viajes y estancias en diferentes países extranjeros le permitieron mantener directa relación con importantes fisiólogos europeos, entre ellos recordemos a Chales Richet e Ivan Petrovich Pavlov.

El relieve histórico de Gómez Ocaña reside principalmente en su actividad docente. Como hemos dicho, enseñó la fisiología en el laboratorio, difundiendo los hábitos de trabajo experimental en un ambiente todavía dominado por acercamientos librescos. Por otra parte, su enseñanza se basaba también en un conocimiento riguroso y al día de la bibliografía especializada en francés, alemán, inglés e italiano.

Como afirma Ruiz de Galarreta, Gómez Ocaña fue un "infatigable repetidor de experimentos y observaciones ajenas". Ello hizo posible que sus monografías sobre la fisiología circulatoria, cerebral y tiroidea no fueran meros resúmenes didácticos, sino excelentes revisiones apoyadas en experimentos propios confirmatorios. Hay que subrayar especialmente que la parte general de sus *Nuevas investigaciones sobre el tiroides* (1895) fue una de las más tempranas exposiciones sistemáticas de la endocrinología. Su punto de partida fue el siguiente: "En tres órdenes de hechos se sustenta la prueba de la secreción del tiroides y de su papel en las funciones nutritivas.

1.º En los efectos que se siguen a la ablación del tiroides en los animales.

2.º En las enfermedades que ofrece el hombre cuando la glándula degenera o se atrofia, y en las consecuencias de la operación del bocio (tiroidectomía).

3.º En los beneficios que resultan del injerto de la glándula o de su administración a los enfermos que sufren los accidentes consecutivos a la lesión o pérdida del tiroides.

En todos los casos, la degeneración o pérdida del tiroides produce en los animales dos clases de accidentes: agudos, que inician la escena patológica y que se caracterizan por convulsiones, disnea y elevación de temperatura (caquexia estrumosa); y crónicos o caquécticos, notables por el apagamiento de todas las funciones, singularmente de las intelectuales, baja de temperatura, paresia o debilidad muscular, suspensión del desarrollo y una hinchazón de la piel que ha dado nombre a la caquexia (mixedema). La muerte es el término inevitable de las dos clases de accidentes, como no se acuda a la medicación por la substancia de la misma glándula.

Parecido carácter tuvo su tratado de fisiología, cuyas sucesivas ediciones fueron aumentando de extensión e incorporando las novedades más recientes. Resulta notable que un experimentador tan entusiasta como Gómez Ocaña intentara conciliar, en algunas cuestiones teóricas, las ideas de Claude Bernard con las de José de Letamendi, por el que tuvo gran admiración personal.

Las aportaciones originales de Gómez Ocaña fueron, por el contrario, limitadas. Las más notables fueron algunas precisiones acerca de la localización de los centros ópticos en la corteza cerebral. Las presentó por vez primera en el XI Congreso Médico Internacional, celebrado en Roma (1894), y las desarrolló en varios trabajos posteriores.

En 1896 las resumió en los siguientes términos: "Ateniéndome al resultado de mis experimentos, me decido a localizar en los hemisferios cerebrales dos clases de centros o aparatos ópticos. Los primeros, o de proyección (L.O., fig. 1), se encuentran en los lóbulos occipitales; reciben directamente las fibras ópticas en la disposición que traen desde su entrecruzamiento parcial en el quiasma, y pueden considerarse como retinas del cerebro, porque en ellas van a proyectarse las imágenes actuales de los objetos presentes. Los segundos, o de representación (P.C., figura 1), se hallan en la región parietal (pliegue curvo en el hombre); reciben de los anteriores las imágenes visuales y las conservan para ser evocadas por la memoria o por la imaginación... Creo, pues, que las regiones parietales reciben fibras del lóbulo occipital de su lado y del opuesto y que, de esta suerte, cada centro óptico de proyección preside a los tres cuartos internos de la retina de nombre contrario y al cuarto externo de la del mismo nombre; y cada centro de representación, a la visión representativa total del ojo opuesto". Amplia repercusión tuvo, por otro lado, una comunicación suya al XIII Congreso Médico Internacional (Madrid, 1903) relativa a la supervivencia de animales afectos de doble vagotomía. Seguida de una demostración en el laboratorio, interesó vivamente a Pavlov y a otros congresistas e influyó más tarde en la investigación de las funciones de los nervios vagos.

Durante sus últimos años, dedicó sus trabajos experimentales a las acciones de los iones calcio, sodio, magnesio y potasio. Algunos de sus resultados, que expuso a partir de 1912 en varios países a través de conferencias y congresos, fueron apreciados internacionalmente. Llegó, incluso, a redactar el artículo sobre el magnesio en el gran *Dictionnaire de physiologie* en diez volúmenes (1895-1923), dirigido por Charles Richet. Basó estos estudios en la observación y registro de fenómenos como la contracción cardíaca o la aparición de convulsiones, ideando con este motivo un nuevo modelo de cardiógrafo y otro de miógrafo. Aunque fue consciente de que se iniciaba una era bioquímica, no llegó a fundamentar sus investigaciones en análisis químicos.

Gómez Ocaña ocupó el cargo de senador desde 1914 hasta su muerte en 1919. Tuvo siempre un gran interés por la historia, el arte y la literatura, siendo, en especial, un estudioso de la obra de Cervantes.

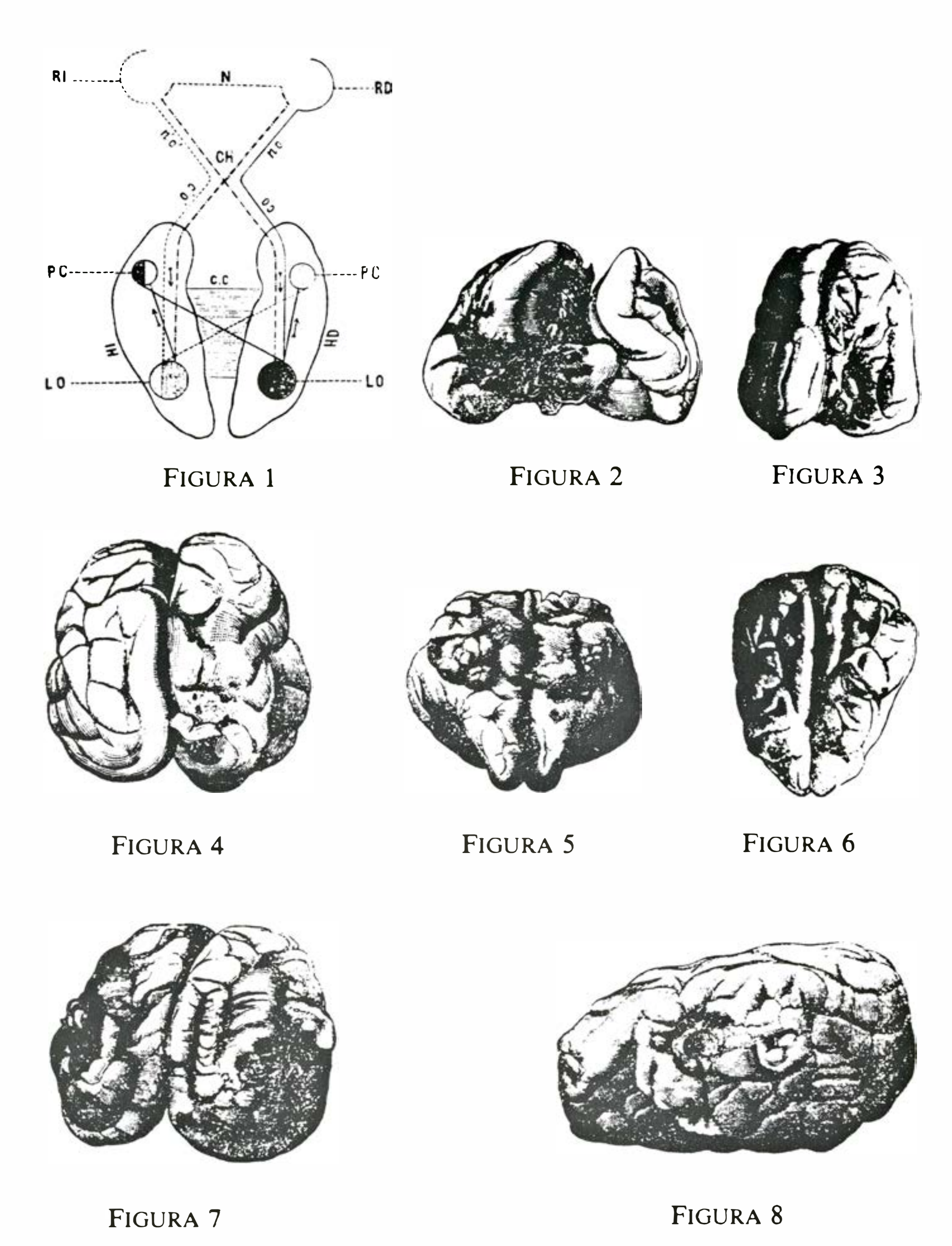


FIGURA 1 FIGURA 2 FIGURA 3

FIGURA 4 FIGURA 5 FIGURA 6

FIGURA 7 FIGURA 8

Figura 1: *Esquema del aparato visual RD y RI, retinas derecha e izquierda; N, porción nasal de ambas retinas; HD y HI, silueta de la porción post-silviana de ambos hemisferios; PC P'C', regiones del pliegue curvo; LO y L'O', lóbulos occipitales; CH, quiasma óptico; CC, cuerpo calloso;* no *y* n'o', *nervios ópticos; y* co *y* c'o', *cintas ópticas. El esquema tiene por objeto mostrar las relaciones de las dos mitades externa e interna de cada retina con los centros cerebrales. A este fin se distingue la porción temporal o externa de la retina izquierda por una línea de puntos; la nasal del mismo lado por puntos y rayas; la temporal de rayas más gruesas. Con estos datos puede el lector abarcar de una sola ojeada la total geografía del aparato óptico. Nótese cómo los centros occipitales presiden la sensibilidad de los tres cuartos internos de la retina opuesta y la del cuarto externo del mismo lado; cómo los centros del pliegue curvo rigen justamente la visión completa del ojo opuesto; y, en fin, cómo el cruce de las fibras en el cuerpo calloso es complementario del que tiene lugar en el quiasma.* **Figura 2:** *Cerebro de gato visto por detrás. La doble lesión causó ceguera completa; la del lado derecho, hemioplía de ambos ojos (de fotografía del natural).* **Figura 3:** *Cerebro de perro. La lesión produjo ambliopía del ojo izquierdo.* **Figura 4:** *Cerebro de gato visto por encima. La lesión del hemisferio derecho causó ceguera del ojo izquierdo.* **Figura 5:** *Cerebro de conejo visto por detrás. La doble lesión ocasionó ceguera completa.* **Figura 6:** *Cerebro de gato visto por detrás. La doble lesión produjo ceguera completa.* **Figura 7:** *Cerebro de perro visto por encima. La doble lesión produjo ceguera completa.* **Figura 8:** *Cerebro de perro visto de lado. La lesión produjo ligeros trastornos visuales.*

El sistema estadounidense de Seguridad Social

La viabilidad del sistema basado en el pago de los gastos con los recursos del momento se ve amenazada por las tendencias económicas y demográficas. Las clases pasivas y las prestaciones aumentan más deprisa que la fuerza laboral activa

Eli Ginzberg

El tercer día de cada mes, unos 35 millones de cheques verdes del gobierno de los Estados Unidos se reparten por las casas o se depositan directamente en los bancos en todo el territorio nacional. Los cheques, cuyo montante medio fue de casi 300 dólares en 1979, representan los pagos mensuales del más importante sistema de transferencias del país: el sistema de la Seguridad Social. Las pensiones totales pagadas a 22 millones de trabajadores jubilados o incapacitados para el trabajo y a 13 millones de dependientes y supervivientes sumaron más de 104.000 millones de dólares en 1979. A diferencia de los sistemas de fondo de pensiones, en los cuales las cotizaciones se utilizan para pagar a los futuros beneficiarios, la Seguridad Social estadounidense funciona a base de ir pagando los gastos presentes con los recursos del momento. Lo cual significa que las cotizaciones pagadas por empleados y empresarios proporcionan los fondos destinados a cubrir prestaciones del momento a los beneficiarios. Estos pagos representan más del 6 por ciento de la renta total disponible de las personas físicas en los Estados Unidos.

Las transferencias de renta a través de la Seguridad Social representan el cumplimiento de un contrato implícito entre las personas más jóvenes, que en estos momentos integran la fuerza de trabajo activa, y las personas que ya han dejado de rendir. Este contrato intergeneracional ha estado en vigor desde que, en 1940, se jubiló la primera persona con derecho a percibir una pensión de la Seguridad Social. Se empieza ahora a dudar de la solidez financiera del contrato. De hecho, durante buena parte del pasado año la nación llegó a creer que el sistema de la Seguridad Social corría el riesgo de una inminente bancarrota. En realidad no era así, pero de no haber mediado las medidas preventivas adoptadas por el Congreso el sistema probablemente habría sufrido un grave problema de *cash-flow* en el plazo de un año. (*Cash-flow* es un anglicismo admitido en economía que designa beneficio más amortización.) Sin embargo, este problema de liquidez traslucía un desequilibrio subyacente entre los ingresos y los gastos. La eficacia del remedio aplicado para resolver las dificultades inmediatas no debería hacernos olvidar las causas originales del desequilibrio y la necesidad de proceder a una reforma en profundidad del sistema a largo plazo.

Los problemas tienen un origen bastante evidente. Puesto que la Seguridad Social es un sistema basado en el pago de los gastos con los recursos del momento, las cotizaciones pagadas por empleados y empresarios en 1950, pongamos por caso, sirvieron para abonar las pensiones percibidas por los beneficiarios en 1950 y las cotizaciones pagadas el presente año servirán para hacer frente a las pensiones de este año. Un sistema de este tipo se halla a merced de las fuerzas económicas y demográficas. Durante largo tiempo estas fuerzas fueron favorables al sistema. En el período de expansión de la economía estadounidense, el empleo creció y los salarios aumentaron al menos a la par con la inflación; una alta tasa de natalidad (que alcanzó su nivel máximo con la explosión de nacimientos de la década de 1950) y un incremento del empleo femenino incorporaron a la fuerza de trabajo muchos más trabajadores de los que ésta iba perdiendo a través de la jubilación. En consecuencia, los ingresos derivados de las cotizaciones o impuestos sobre la nómina de salarios fueron superiores a las pensiones pagadas.

La situación está cambiando. Los salarios (que, junto con el nivel de empleo y el tipo impositivo, determinan el nivel de ingresos derivados del impuesto sobre la nómina de salarios) no han aumentado a la par con la inflación, mientras que las pensiones (que van ligadas al nivel de precios) se hallan protegidas, tal vez incluso en exceso. Y, lo que es más importante a largo plazo, una menor fertilidad y una mayor longevidad están modificando la estructura de edad de la población estadounidense, con un incremento del número de beneficiarios potenciales en relación al número de trabajadores que deberán sufragar sus gastos. Una persona de 18 años que empiece a trabajar actualmente en un empleo cubierto por la Seguridad Social no estará en edad de jubilarse hasta la década de 2020; el sistema puede verse obligado a pagarle pensiones, a él y a su viuda e hijos, hasta la década de 2040. En palabras de un analista, un 85 por ciento de las personas que cobrarán pensiones durante los próximos 75 años viven ya en la actualidad; un 96 por ciento de todas las pensiones que pagará la Seguridad Social durante los próximos 75 años serán percibidas por personas que viven ya en la actualidad; un 99 por ciento de los impuestos sobre la nómina de salarios que se pagarán en los próximos 25 años y un 81 por ciento de los impuestos que se pagarán en los próximos 50 años serán sufragados por personas que ya andan por la calle.

El Congreso no tiene más remedio que plantearse cómo hará frente a los compromisos que está adquiriendo en la actualidad. Los vástagos de la explosión demográfica empezarán a jubilarse antes del año 2010. Disponemos por tanto de unos 30 años de plazo para reestructurar el sistema de la Seguridad Social.

En los Estados Unidos se introdujo tardíamente la idea de que las personas debían esperar del gobierno central una protección contra la pérdida de ingresos por motivo de vejez o invalidez y la atención de las necesidades económicas de las personas que quedan desatendidas en caso de fallecimiento del cónyuge o padres. Alemania ya había establecido un programa de pensiones bajo el patrocinio del gobierno en 1889, el Reino Unido en 1908 y Francia en 1910. La ley sobre la Seguridad Social se aprobó en 1935, durante el apogeo del New Deal. El presidente Roosevelt (desoyendo las objeciones de algunos de sus asesores) decidió restringir inicialmente las prestaciones a los trabajadores industriales y del comercio y limitar éstas al pago de una pensión monetaria a partir del momento en que dejaran de trabajar, a los 65 años de edad; una edad de jubilación fijada sin la suficiente reflexión. Sin embargo, se daba por sentado que más adelante se ampliaría el sistema. Ya antes de pagarse las primeras prestaciones, el Congreso introdujo una enmienda en la ley que contemplaba el pago de pensiones a la esposa anciana y a los hijos de un trabajador jubilado y a la viuda e hijos de un trabajador asegurado.

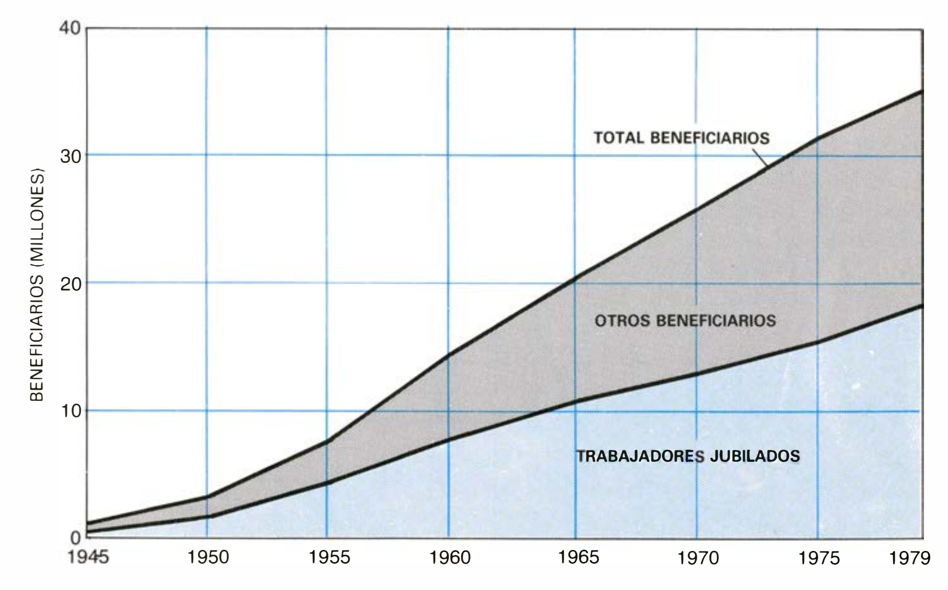
Posteriormente se acoplaron al sistema original otros dos grandes programas. En 1956 se añadió el seguro de invalidez para los trabajadores asegurados que quedaran incapacitados para el trabajo por un accidente o enfermedad. En 1965 se añadió el seguro médico, Medicare. Este se divide en dos partes: seguro hospitalario y seguro médico complementario, que contribuye a sufragar los honorarios médicos que debe pagar el paciente. El núcleo original del sistema de Seguridad Social se denomina formalmente Seguro para la Vejez y los Supervivientes (*Old Age and Survivors' Insurance*, OASI). Cuando se incluye el seguro de invalidez, el programa se llama OASDI, y, con el seguro hospitalario, OASDHI. Todos estos programas se financian a partir de los impuestos sobre la nómina de salarios. El seguro médico complementario se alimenta de fondos independientes y es un programa opcional, aunque la mayoría de los beneficiarios del seguro Medicare optan por participar en él; las primas pagadas por los beneficiarios cubren actualmente el 30 por ciento de los costes y el resto se sufraga a cargo de los ingresos generales del Gobierno Federal.

El sistema cubre actualmente alrededor del 90 por ciento de los asalariados. Estos deben pagar un impuesto sobre la nómina de salarios (cifrado en un 6,7 por ciento sobre los salarios de hasta 32.400 dólares) que tiene como contrapartida un impuesto equivalente pagado por los empresarios. Una importante proporción de los más de 8 millones de trabajadores autónomos están acogidos también al sistema y pagan un 75 por ciento del impuesto conjunto pagado por el empleado y la empresa. La categoría más numerosa de trabajadores que no se hallan cubiertos por la Seguridad Social la componen los empleados de la administración pública, quienes cuentan con 68 planes federales de jubilación y millares de otros planes para los empleados de la administración local y de los estados federales. Tampoco se hallan cubiertos los trabajadores de algunas instituciones no lucrativas.

Desde sus inicios, el sistema de la Seguridad Social se planteó con vistas a cubrir dos objetivos contradictorios. La intención fundamental era crear un sistema obligatorio de cotizaciones forzosas que servirían para proporcionar unos ingresos después de la jubilación. De acuerdo con la predilección nacional por el sistema de "escoger cada uno su propia cantidad a pagar" y en vistas a obtener un amplio apoyo de la opinión pública, se vincularon las prestaciones al montante de las "aportaciones". Cuanto más elevadas fueran las cotizaciones pagadas por el trabajador, mayor sería la pensión que percibiría al jubilarse.

Pero la segunda finalidad era asegurar, también a las personas que habían percibido unos salarios bajos, unos ingresos suficientes para cubrir sus necesidades básicas en la vejez. Ello no podía lograrse con unas pensiones directamente proporcionales a las cotizaciones pagadas; la relación pensión-cotización tenía que alterarse para favorecer a los perceptores de salarios bajos. Actualmente, el trabajador varón soltero medio que se jubiló en 1979 después de percibir salarios equivalentes a la máxima categoría imponible habrá percibido en el momento de su muerte prestaciones equivalentes a 4,97 veces lo que pagó en concepto de cotizaciones; para las personas solteras con un salario bajo, la relación prestaciones totales-cotizaciones totales es de 7,07, para los hombres, y 8,82 para las mujeres. Para una persona jubilada con un cónyuge dependiente esta relación aumenta hasta 9,19 para el perceptor del salario máximo y 13,06 para el perceptor del salario mínimo. Evidentemente, estas cifras tan elevadas sólo han sido posibles gracias al sistema de pagar los gastos con los recursos del momento, combinado con las fuerzas demográficas ya citadas. (Por cierto, con frecuencia se señala que el impuesto de la Seguridad Social es regresivo, pues no sólo se aplica el mismo tipo impositivo a todos los niveles de ingresos sino que, además, la fracción de los ingresos a la que se aplica el impuesto es mayor en el caso de una persona con un nivel bajo de ingresos que en el de una persona con una renta alta. Sin embargo, desde el punto de vista de las prestaciones, el sistema es claramente progresista, puesto que favorece a las personas con un bajo nivel de ingresos.)

La pensión que percibe una persona



NUMERO DE BENEFICIARIOS del Seguro para la Vejez, los Supervivientes y la Invalidez (OASDI). Ha crecido sin cesar desde que se jubiló el primer trabajador asegurado, en 1940. El seguro acoge a trabajadores jubilados, incapacitados para el trabajo, cónyuges e hijos dependientes y supervivientes.

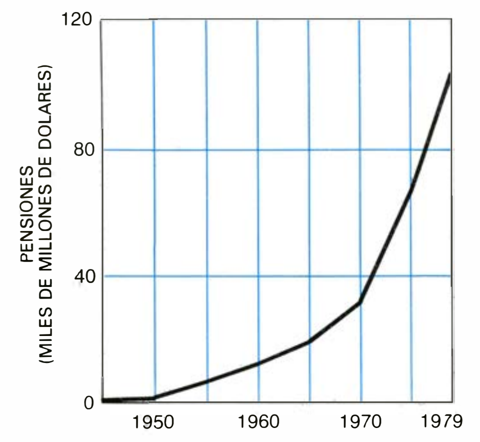
al jubilarse se basa en sus ingresos obtenidos de los empleos sometidos al impuesto sobre la nómina de salarios; aunque durante algunos años estos ingresos pudieran haber sido superiores a la cantidad máxima imponible, se computa únicamente esta cifra máxima. Para calcular el montante de las pensiones primero se ponderan los ingresos de acuerdo con el aumento habido en los niveles de salarios desde que se percibió ese dinero. Después de excluir algunos años de ingresos mínimos se dividen por 12 los ingresos anuales ponderados percibidos hasta los 62 años de edad y se efectúa el promedio. A continuación se aplica a los ingresos mensuales medios ponderados una fórmula que permite determinar el "montante primario del seguro" (*"primary insurance amount"*, PIA). La fórmula es compleja, pero en esencia lo que hace es convertir los incrementos progresivamente más elevados de ingresos en el "montante primario", equivalente a porcentajes progresivamente más bajos. En consecuencia, la pensión a percibir por un trabajador con un nivel bajo de ingresos representa un porcentaje mayor de su salario medio que la pensión a percibir por un trabajador con un alto nivel de ingresos. Este porcentaje global se denomina tasa de restitución. Oscila entre casi el 70 por ciento para los trabajadores con el salario más bajo hasta alrededor de un 33 por ciento para quienes han percibido cada año el salario máximo imponible; la tasa media de restitución es de un 40 a un 50 por ciento. El "montante primario" representa la pensión que cobrará un trabajador jubilado de 65 años sin cargas familiares. Esta cantidad se pondera si la persona se jubila antes o después de los 65 años o si tiene un cónyuge o hijos dependientes.

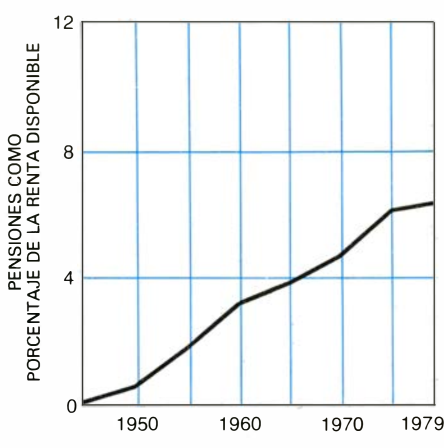
En la medida en que la Seguridad Social se considera un plan social de seguro, las personas que cotizan tienen derecho a recibir prestaciones. Esta noción de derecho adquirido ha protegido a la Seguridad Social de los ataques políticos (que ya había previsto el presidente Roosevelt) y también ha actuado como elemento disuasorio ante cualquier intento de reducción de las prestaciones una vez concedidas. (No obstante, yo personalmente pienso que la elevada relación prestaciones-cotizaciones pagadas contradice la idea de que cualquier reducción de las prestaciones representaría un abuso de confianza con respecto a las personas que han estado pagando el impuesto, en particular si la modificación se anuncia con la debida antelación.)

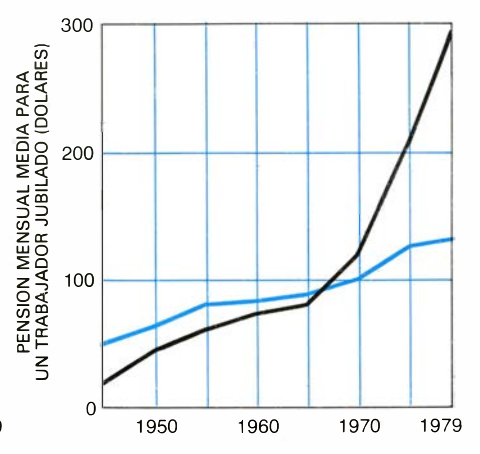
La concepción del sistema como seguro basado en el pago de cuotas se ha visto reforzada por la decisión de sufragar todas las prestaciones del OASDHI a partir del impuesto sobre la nómina de salarios. La mayoría de los planificadores del sistema partían de la base de que con el tiempo éste se financiaría en parte con los ingresos generales del Estado. Pensaban en una posterior división tripartita, característica de la mayoría de sistemas de pensiones europeos, en virtud de la cual, empleados, empresarios y gobierno central aportan, respectivamente, una tercera parte de los fondos necesarios. El nexo directo entre cotizaciones y prestaciones ha ayudado al Congreso a hacer frente a las presiones por un incremento de las prestaciones, sin tener en cuenta su fuente de financiación.

Es imposible analizar el tema de la Seguridad Social sin traer a colación otras medidas encaminadas a reforzar el nivel de ingresos. Las mismas razones que indujeron al gobierno federal a promulgar la Ley de la Seguridad Social condujeron a la creación de diversos programas encaminados a proporcionar ayudas en dinero y otras prestaciones a aquellas personas con unos ingresos demasiado bajos para llegar a cubrir sus necesidades básicas. En tanto que las prestaciones de la Seguridad Social se pagan a todos los trabajadores con derecho a ellas y que han cotizado sobre sus ingresos, independientemente de que las necesiten o no, casi todos los demás programas dependen de los recursos de que se dispongan y sólo pueden beneficiarse de ellos quienes demuestren hallarse en situación de indigencia.

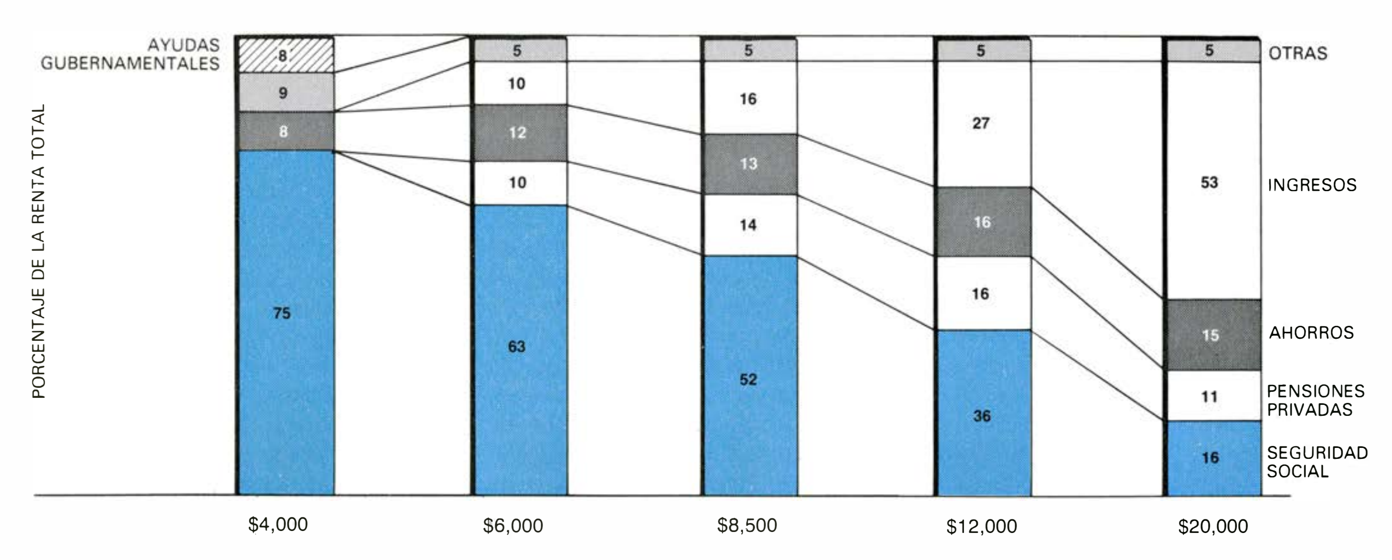
Los programas sometidos a la inspección de recursos se iniciaron modestamente en la década de 1930 con un programa conjunto del gobierno federal y los gobiernos de los estados destinado a prestar ayuda a las familias con hijos dependientes (*Aid to Families with Dependent Children*, AFDC), que todavía representa la partida principal de los llamados "pagos de beneficencia". También se prestó ayuda a los ciegos y, más adelante, a las personas inválidas y ancianos necesitados que no estaban cubiertos por la Seguridad Social o recibían pensiones insuficientes; a principios de la década de 1970 el gobierno federal se hizo cargo de estos grupos con el programa de ingresos supletorios (*Supplemental Security Income*). En 1965 se creó el programa de asistencia médica, Medicaid, financiado conjuntamente por el gobierno federal y los distintos estados y destinado a las personas que recibían prestaciones benéficas y a todas aquellas calificadas, en cada estado, como médicamente necesitadas. Otros programas de ayuda no monetaria están encaminados a ofrecer alimentación (cupones para la adquisición de alimentos [*food stamps*] y otras ayudas) y vivienda (alquileres subven-







IMPORTE ANUAL de las pensiones pagadas por el OASDI; ha aumentado de año en año en términos monetarios y también como proporción de la renta total disponible en los Estados Unidos. En el gráfico puede observarse el efecto de los aumentos decretados durante los años 1970. También ha aumentado la pensión mensual media de un jubilado, pero el valor de la pensión ajustado a la inflación (en dólares de 1967) ha aumentado más lentamente (*color*).



IMPORTANCIA de la Seguridad Social como fuente de ingresos después de la jubilación. Se indica para cinco niveles de renta de 1978. Los programas gubernamentales de ayuda sometida a la inspección de recursos aportaron el 8 por ciento de los ingresos de las familias con una renta de 4000 dólares anuales.

cionados) a las personas con un bajo nivel de ingresos. En 1979, el gasto total del gobierno federal en programas sometidos a la inspección de recursos fue de 35.000 millones de dólares, frente al gasto total de 104.000 millones de dólares en el OASDI.

Pese al propósito general de lograr que la Seguridad Social cubra las necesidades básicas de las personas incapacitadas para el trabajo por razones de edad o invalidez, más de dos millones de beneficiarios del OASDI tuvieron que recibir ingresos supletorios en 1979. Si sumamos todos los ingresos, incluidas las prestaciones, monetarias y no monetarias, de la Seguridad Social y de los programas sometidos a la inspección de recursos, las personas de edad que quedan por debajo del límite de pobreza no sobrepasan el 3 o 4 por ciento del total. Pero una tercera parte de todas las personas de edad se encuentran apenas por encima del límite. Evidentemente, nunca se pensó que las pensiones de la Seguridad Social pudieran constituir la única fuente de ingresos para las personas de edad. La mayoría de las familias de edad avanzada que gozan de un nivel de vida satisfactorio reciben menos de la mitad de sus ingresos en forma de pensiones de la Seguridad Social; también cuentan, en diversas proporciones según los distintos niveles de renta, con ingresos de otros miembros de la familia, con sus ahorros y con pensiones privadas.

Los sistemas privados de pensiones son anteriores a la creación de la Seguridad Social, aunque sólo adquirieron un peso significativo durante y después de la segunda guerra mundial. De 1950 a 1975 el número de trabajadores cubiertos por seguros privados pasó de 9,8 a 30,3 millones, mientras que el número mensual de beneficiarios pasaba de 450.000 a más de siete millones y las prestaciones pagadas anualmente aumentaban de 370 millones a 14.800 millones de dólares. En 1979 casi dos terceras partes de todos los trabajadores varones y la mitad de todas las trabajadoras estaban cubiertos por un seguro privado. No obstante, en 1980 esos seguros representaban sólo el 14 por ciento de todas las pensiones de jubilación, invalidez, viudedad y orfandad; un 78 por ciento correspondía a la Seguridad Social y otros programas federales de jubilación y un 8 por ciento correspondía a programas estatales y locales.

Queda bien patente la importancia fundamental del sistema de la Seguridad Social. Se ha dicho con frecuencia (y probablemente sea cierto) que constituye el programa social del gobierno que mejores resultados ha dado. Lo cual hace aún más imperioso mantener financieramente saneado el sistema.

La crisis de liquidez (*cash-flow*) que se cernía amenazante el pasado otoño sólo afectó a una parte del sistema de la Seguridad Social (OASDHI). La legislación vigente establece que las prestaciones correspondientes a cada programa de la Seguridad Social se paguen únicamente con los fondos fiduciarios destinados a ese programa en concreto. Y el que está en dificultades de fondos es el programa OASI (seguro de vejez y pensiones a los supervivientes). A principios de 1960, el dinero depositado en ese fondo representaba el 195 por ciento del gasto total para 1960; el pasado primero de octubre (inicio del año fiscal de 1982) el saldo representaba sólo el 14 por ciento del gasto previsto para 1982. La Oficina de Presupuestos del Congreso se guía por esta norma indicativa: el saldo inicial al principio de cada año debe representar al menos el 9 por ciento del gasto previsto para ese año, a fin de poder cubrir las prestaciones que se pagan al principio del año fiscal, antes de recaudarse el impuesto sobre la nómina de salarios para ese año. De no haber mediado la intervención del Congreso, el saldo de la OASI habría sido inferior al 5 por ciento, en octubre de 1982.

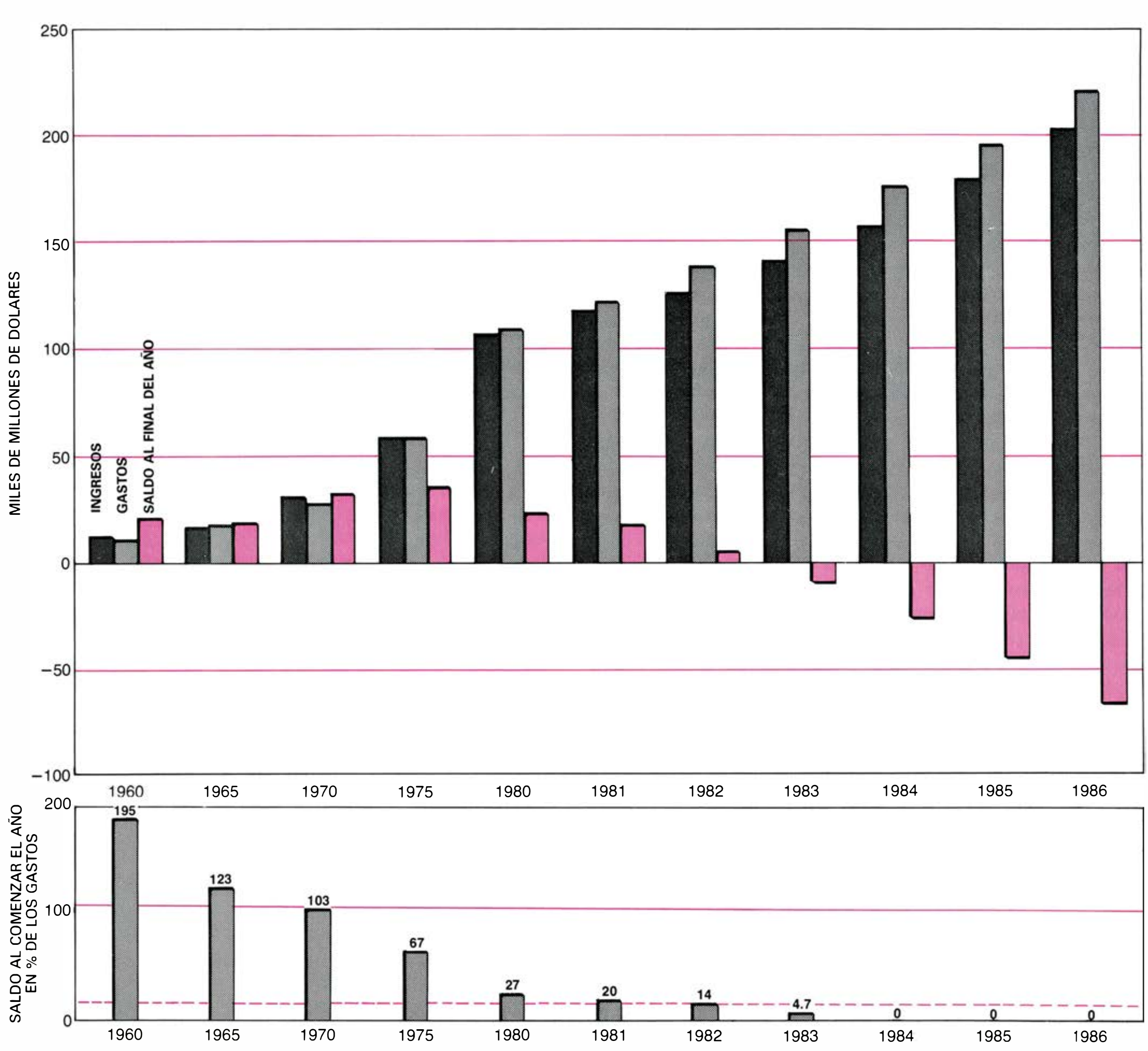
El motivo fundamental de la disminución de los saldos del fondo fiduciario durante la década de 1970 fue una sucesión de tres grandes incrementos de las pensiones: de un 15 por ciento en 1970, un 10 en 1971 y un 20 en 1972. Además, en 1972 se institucionalizó el incremento continuado de las prestaciones. En efecto, el Congreso decidió que las pensiones aumentasen automáticamente para compensar cualquier incremento superior al 3 por ciento anual en el índice de precios al consumo. En consecuencia, las pensiones han venido aumentando cada año desde 1975. Los incrementos debían financiarse aumentando la base salarial imponible a la par con el aumento de los salarios. Pero la subida de los precios fue superior a la de los salarios y las pensiones reajustadas aumentaron más rápidamente que la base salarial reajustada.

En 1977, el Congreso, obligado a hacer frente a los efectos combinados del aumento de las pensiones, el elevado nivel de desempleo y el escaso crecimiento económico (y sin ningún deseo de modificar la financiación basada en el impuesto sobre la nómina de salarios y aún menos inclinado a reducir las pensiones), decretó una serie de incre-

mentos en el tipo total de cotización a la Seguridad Social (OASDHI). Se esperaba que este aumento del tipo impositivo, unido al incremento de la base imponible, lograría asegurar la estabilidad financiera del sistema durante lo que quedaba de siglo. Pero no fue suficiente. Pasados menos de tres años, el Congreso tuvo que transferir algunos ingresos tributarios del fondo del seguro de invalidez al fondo de pensiones de la OASI. Y el pasado otoño viose obligado a adoptar nuevas medidas de emergencia, traspasando algunos futuros ingresos tributarios de los fondos del seguro de invalidez y hospitalario al fondo de pensiones de la OASI y autorizando también posibles créditos interfondos. Sin embargo, de las proyecciones realizadas por la Oficina de Presupuestos del Congreso se desprende que ni los créditos interfondos ni tampoco una completa fusión de los distintos fondos fiduciarios permitirá subsistir durante largo tiempo al sistema. Cuando se inicie el año fiscal 1985, el saldo conjunto de los tres fondos será inferior al 9 por ciento de los gastos previstos.

¿Qué hacer entonces? En febrero de 1981, la Oficina de Presupuestos del Congreso expuso una serie de alternativas que, con diversas combinaciones, podrían permitir la supervivencia del sistema hasta finales de la década de 1980. Por ejemplo, muchos economistas opinan actualmente que los incrementos del coste de la vida tienden a sobrecompensar la inflación y que podrían modificarse. Se podría suprimir la pensión mínima, que incrementa el pago que de otro modo recibirían algunas personas con salarios muy bajos. El Congreso ya ha suprimido la pensión "en bloque" en caso de fallecimiento y las ayudas especiales para los estudiantes dependientes.

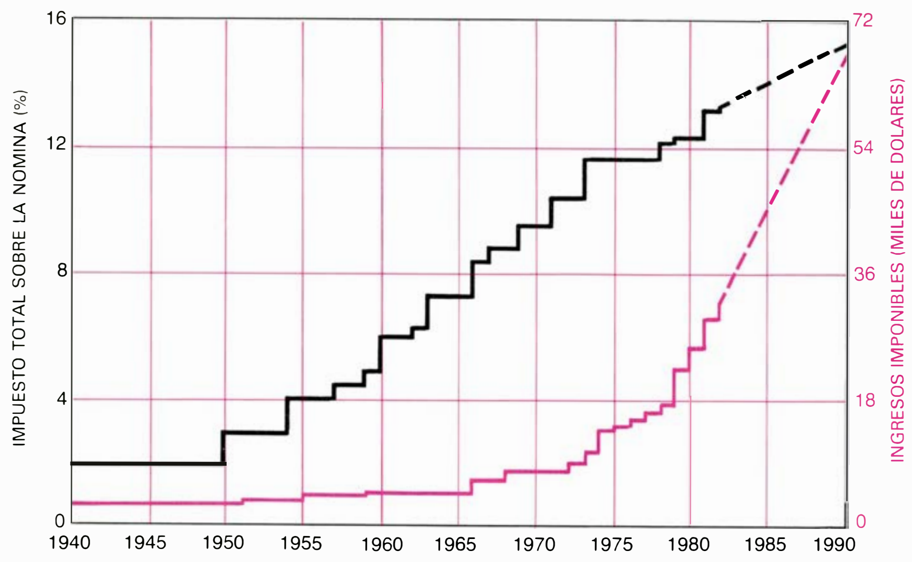
La Oficina de Presupuestos ha sugerido también una serie de procedimientos para aumentar los ingresos, como por ejemplo autorizar la concesión de créditos del erario público al sistema, la aceleración de los aumentos programados en las cotizaciones, el aumento de la base imponible incluyendo en ella to-



FONDO FIDUCIARIO de la OASI, con el que se pagan las pensiones a los trabajadores jubilados y sus dependientes y supervivientes ("las cargas familiares"). Estos últimos años ha gastado anualmente más de lo que ha ingresado en concepto de cotizaciones. El cuadro superior muestra el progresivo incremento de los gastos, que han superado a los ingresos, reduciendo el saldo final de cada año y creando una expectativa de saldos negativos para mediados de nuestra década. El cuadro inferior muestra el saldo del fondo al comienzo de cada año fiscal expresado como porcentaje del gasto de ese año. (Hasta 1980 los porcentajes se basan en el gasto real; desde 1981 a 1986, en estimaciones de la oficina de Presupuestos del Congreso.) Se considera esencial contar con un saldo inicial de un 9 por ciento para que el sistema haga frente a sus obligaciones antes de recaudar las cotizaciones correspondientes al ejercicio. El saldo del próximo año habría quedado por debajo del límite (*línea de color a trazos*) de no haber intervenido el Congreso el pasado otoño.

dos los ingresos y el aumento de la cotización de los trabajadores autónomos. Las prestaciones hospitalarias podrían financiarse totalmente o en parte con dinero procedente de ciertas partidas específicas del impuesto sobre la renta, con lo cual toda la partida hospitalaria del impuesto sobre la nómina de salarios podría destinarse al pago de pensiones de vejez, a dependientes e invalidez (OASDI). El Congreso podría considerar la posibilidad de prescindir de algunas futuras reducciones del impuesto sobre la renta y asignar en cambio una parte de cualquier reducción propuesta a los fondos fiduciarios de la Seguridad Social. Otros han sugerido que la mitad de cualquier pensión pagada debería estar sometida al impuesto sobre la renta; concretamente, la mitad financiada con las aportaciones de los empresarios, sobre la que no se ha pagado nunca ningún impuesto. Aunque los beneficiarios con un bajo nivel de ingresos estarían exentos del pago del impuesto, la recaudación representaría unos 4500 millones de dólares al año.

Con tantas opciones parece sumamente improbable que el sistema no pueda cumplir sus compromisos en los próximos años. Un aumento de sólo un 0,5 por ciento en el impuesto total sobre la nómina de salarios, unido a las medidas ya adoptadas por el Congreso, restablecería el equilibrio entre ingresos y gastos a menos que la economía entrase en un prolongado declive.



TIPO DEL IMPUESTO sobre la nómina de salarios (*negro*) y los ingresos máximos, o base imponible, sobre los que se paga el impuesto (*color*); han aumentado persistentemente y continuarán haciéndolo. Los futuros incrementos del tipo impositivo se hallan especificados legalmente. Sin embargo, desde principios de 1970, la base máxima imponible se reajusta de acuerdo con el nivel medio nacional de salarios; los incrementos futuros representados son estimaciones. Las ilustraciones del artículo son de Ilil Arbel.

El verdadero desafío al que deberá enfrentarse la Seguridad Social se sitúa en un futuro más lejano. La proporción de la población estadounidense formada por personas de 65 años y más ha aumentado (de un 9,2 por ciento en 1960 a un 11,2 el pasado año) y continuará creciendo (hasta alcanzar un porcentaje estimado en el 12,2 en el año 2000). Los jubilados y sus cónyuges viven más tiempo que antes: en 1950 la esperanza media de vida a los 65 años era de otros 13 años para los varones y 15 años para las mujeres, mientras que en 1980 se alcanzaron los 15 años para los varones y los 20 años para las mujeres. El factor clave que determina la integridad financiera del sistema de la Seguridad Social es la relación entre el número de trabajadores que cotizan y el número de trabajadores jubilados, personas dependientes y supervivientes con derecho a recibir prestaciones. En 1945, los pagos corrientes a cada beneficiario contaban con las cotizaciones de 50 trabajadores para financiarse; en la actualidad, cada beneficiario depende de las cotizaciones de unos tres trabajadores y en el año 2035 esta relación puede ser de menos de dos contribuyentes por beneficiario.

Los ministros de Asistencia Sanitaria y Social, de Trabajo y de Hacienda, que administran fiduciariamente el sistema de la Seguridad Social, han estudiado las perspectivas a largo plazo. En su informe de 1980 estimaban el saldo medio anual entre ingresos y gastos para los próximos 75 años bajo el supuesto de que no se introdujera ninguna modificación en la base imponible y en los aumentos de la tasa de cotización actualmente prescritos. Las estimaciones se realizaron bajo tres conjuntos de supuestos: optimistas, intermedios y pesimistas. Según el informe, entre 1980 y 2004 el sistema en su conjunto presentará un saldo anual favorable entre ingresos y gastos. Durante el siguiente cuarto de siglo, la proyección intermedia cifra el déficit anual en unos 12.000 millones de dólares y, entre 2030 y 2054, esa proyección cifra el déficit en unos 46.000 millones al año.

El remedio más evidente para un déficit es un aumento de las cotizaciones. La Comisión Nacional sobre la Seguridad Social ya calculó el pasado mes de marzo el incremento de los ingresos obtenidos de las cotizaciones que sería necesario para mantener equilibrado el sistema, suponiendo que el Congreso decrete algunas de las modestas modificaciones en las pensiones que recomienda la Comisión. Ya está previsto que la cotización total (la parte que paga el empresario más la parte que paga el empleado) aumente desde el 13,4 por ciento actual hasta el 15,3 en 1990. A lo cual sería preciso sumar otros incrementos adicionales que subirían el total hasta un 15,7 por ciento entre 2005 y 2009, hasta un 20 por ciento entre 2020 y 2024 y hasta un 23,8 por ciento entre 2035 y 2039. La comisión veía dos riesgos en un incremento del impuesto sobre la nómina de salarios hasta esos niveles: la cantidad a pagar en concepto de cotizaciones podría resultar demasiado importante para los trabajadores jóvenes y otros trabajadores con salarios bajos, y la parte que deberían pagar los empresarios podría frenar la contratación de nuevos trabajadores. La comisión recomendó, por tanto, limitar el impuesto sobre la nómina de salarios a un 18 por ciento y cubrir la diferencia entre esta cifra y el nivel de cotización recomendado con fondos procedentes de los ingresos generales del Estado.

¿Está totalmente descartado que las cotizaciones a la Seguridad Social puedan sufrir un incremento de casi el 80 por ciento? La comisión compartía esta opinión. Los sistemas de Seguridad Social de algunos países de Europa occidental se basan en un tipo de impuesto sobre la nómina de salarios (empleado más empresario) de entre el 18 y el 28 por ciento. A cambio de unas cotizaciones más elevadas, las prestaciones abarcan también un campo más amplio. En esos países, las contribuciones para otros programas sociales públicos y privados pueden hacer subir la carga tributaria total hasta un 50 por ciento de los ingresos de un trabajador.

Antes de continuar examinando la financiación del sistema de la Seguridad Social, debemos preguntarnos si las prestaciones que ofrece actualmente el sistema son adecuadas. Personalmente sostengo que el sistema podría

mejorar y que estas mejoras incrementarían los futuros costes. Entre los problemas más patentes tenemos la insuficiencia de la presente pensión mínima, las discriminaciones hacia las mujeres y la ineficiencia del seguro de asistencia médica, Medicare, para subvenir a las necesidades sanitarias de las personas ancianas.

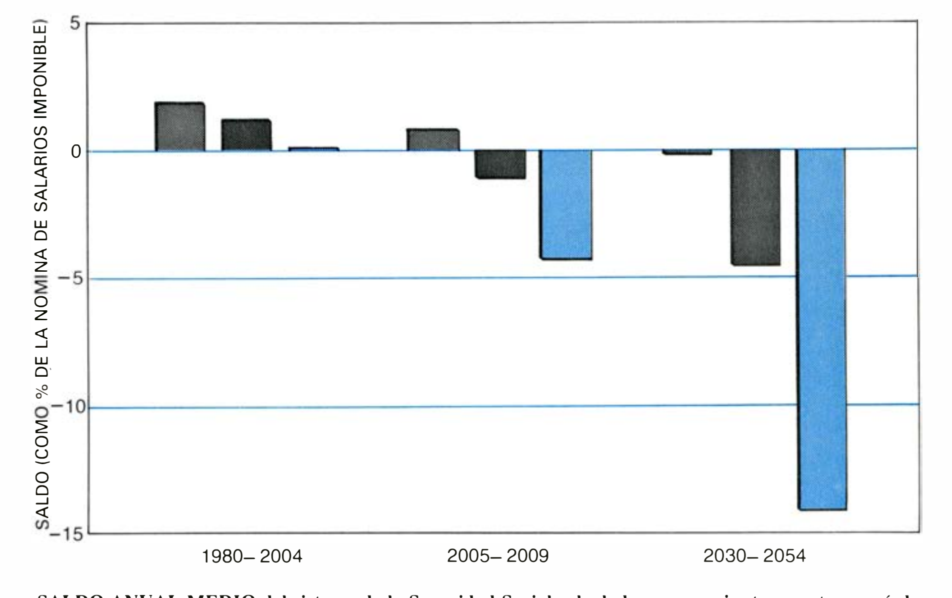
Cualquier trabajador jubilado con derecho a recibir una pensión cobra al menos 122 dólares mensuales. Esta pensión mínima la perciben principalmente mujeres que cobraron salarios muy bajos mientras trabajaban. El Consejo Asesor sobre Financiación y Pensiones de la Seguridad Social declaró, en 1979, que "las personas solteras que han trabajado a jornada completa cobrando el salario mínimo federal no reciben en estos momentos una pensión suficiente para mantenerse fuera de la pobreza". El Consejo recomendaba la adopción de medidas correctoras. El presidente Reagan siguió el curso de acción contrario y recomendó la supresión de la pensión mínima, alegando que es una ganga inmerecida para las personas que se han jubilado con otras pensiones federales pero que también quedan incluidas por un escaso margen entre los beneficiarios de la Seguridad Social. El Congreso se plegó a los deseos del Presidente en la primera ronda de cortes presupuestarios, pero volvió a restablecer la pensión mínima el pasado otoño. Su montante continúa siendo inadecuado.

Cuando se creó el sistema de la Seguridad Social, las mujeres recibían prestaciones principalmente en concepto de dependientes o supervivientes. Esta situación ha cambiado con la entrada de gran número de mujeres en la fuerza de trabajo. En 1978, dos de cada cinco beneficiarias recibían prestaciones por razón de su propio trabajo y no por matrimonio; es probable que esta proporción llegue a ser del 80 por ciento para las mujeres que cumplirán 65 años el año 2000. Las trabajadoras cotizan igual que los varones, pero muchas mujeres cobran pensiones más elevadas en tanto que dependientes que como trabajadoras jubiladas, pues la "tasa de restitución" opera en contra suya. Un resultado desfavorable es que un matrimonio de 65 años cobra un 17 por ciento más en pensiones si sólo uno de ellos ha trabajado como asalariado que si ambos ganaban la misma cantidad. Si una mujer se divorcia antes del décimo año de matrimonio no cobra nada cuando su ex-marido se jubila; si el matrimonio dura 10 años, la mujer cobra la pensión completa del marido.

Una propuesta que podría hacer más justicia a las mujeres es un plan de "ingresos comunes". Las pensiones se basarían en la suma de los ingresos del matrimonio, tanto si han trabajado ambos como si no, e independientemente de la cantidad que cada cual ha ganado y, por tanto, también de cuánto ha cotizado cada uno a la Seguridad Social. En efecto, todos los ingresos se adjudicarían equitativamente al marido y a la mujer. El Consejo Asesor se muestra favorable a la implantación de un plan de este tipo, pero sugiere que se efectúen nuevos estudios para asegurarse de que ello no mermaría las pensiones que perciben actualmente muchas mujeres en tanto que dependientes o supervivientes. De momento, el Consejo recomienda dos posibles medidas; a saber: permitir que el cónyuge superviviente perciba una pensión basada en los ingresos conjuntos del matrimonio e introducir una forma limitada del sistema de "ingresos comunes" para las parejas divorciadas.

El seguro médico Medicare cubre actualmente sólo alrededor de un 40 por ciento de los costes de la atención médica a los ancianos. Cubre sobre todo la atención hospitalaria en casos graves y los honorarios médicos en el caso de las personas que deciden pagar el seguro médico complementario. Cada año un número considerable de personas de edad avanzada incurren en elevadísimos gastos médicos (más de 2500 o 3000 dólares) que deben pagar de su propio bolsillo. El seguro no cubre buena parte de los cuidados necesarios en caso de dolencias crónicas. Medicare no contribuye en absoluto a sufragar los gastos de un paciente crónico o debilitado para que pueda permanecer en su propia casa al cuidado de una enfermera o asistenta domiciliaria y tampoco contribuye de modo significativo a sufragar los gastos de su internamiento en una casa de reposo. Al no responsabilizarse de los gastos de la atención médica del paciente en su propia casa o en una casa de reposo, Medicare alienta la permanencia innecesaria en los hospitales, incrementando con ello los gastos totales de asistencia sanitaria de la nación. Aún así, la cotización al seguro hospitalario aumentaría notablemente a partir de 1990.

Los imperativos demográficos y la necesidad de mejorar las prestaciones de la Seguridad Social plantean varios interrogantes. El primero es: ¿Podrá seguir manteniéndose el sistema hasta mediados del próximo siglo en base a los incrementos del impuesto sobre la nómina de salarios? Sin duda es factible introducir un moderado incremento del tipo impositivo y podría implantarse el pago sobre todos los ingresos. Pero dado que el tipo impositivo es igual para las personas con ingresos bajos que para aquellas con ingresos elevados, la cantidad a pagar pronto resultaría demasiado desproporcionada para algunos asalariados. Y en lo tocante a la parte a pagar por el empresario, no cabe la menor duda de que un impuesto sobre la nómina de salarios demasiado elevado puede inhibir la contratación, creando graves problemas a la economía y la sociedad; puede que esto ya haya empezado a ocurrir en algunos países de Europa occidental. No cabe aumentar indefinidamente el



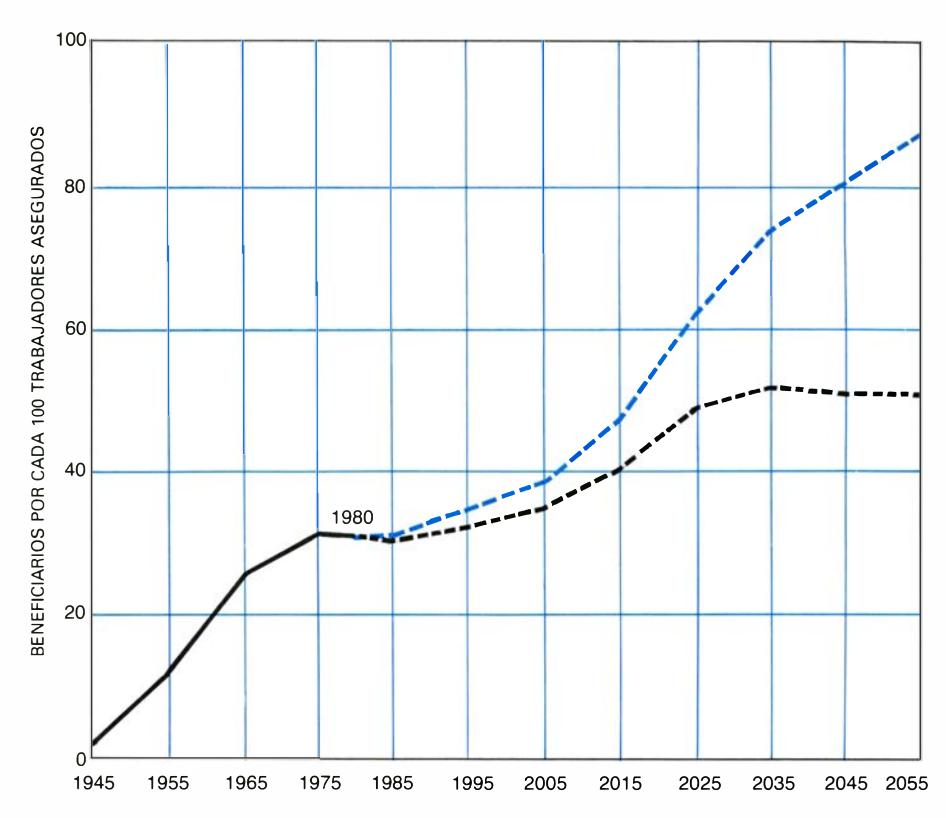
SALDO ANUAL MEDIO del sistema de la Seguridad Social calculado en su conjunto para tres períodos de 25 años bajo supuestos optimistas (*gris*), intermedios (*negro*) y pesimistas (*color*). Las estimaciones (y así se representan aquí) no se hicieron en dólares sino en términos de porcentaje de la nómina total de salarios sometida al pago del impuesto y, por tanto, son válidas independientemente de la inflación. Los porcentajes pueden traducirse a dólares corrientes: un 1 por ciento equivale a 10.000 millones.

tipo impositivo sin consecuencias tan desfavorables como puedan ser la merma de los salarios, la ampliación de la actividad económica incontrolada y una disminución de la recaudación total.

¿Existe algún otro procedimiento para aumentar los ingresos de la Seguridad Social? Una interesante posibilidad, que debería aprovecharse tenazmente, es incorporar al sistema el mayor número posible de empleados públicos y de instituciones no lucrativas que ahora no están integrados en él. Los grupos que no pueden acogerse a la Seguridad Social representan alrededor del 10 por ciento de la población activa. Quizá no pueda obligarse a acogerse al sistema a los trabajadores actualmente cubiertos por planes estatales o locales, pero el Congreso sin duda podría ampliar la cobertura a todos los empleados federales. Sería necesario introducir complicados reajustes para proteger los intereses de estos funcionarios; las dificultades son enormes, pero no insuperables. Muchos expertos en el sistema de la Seguridad Social propugnan la propia implantación de un programa de consolidación de este tipo y argumentan que ningún otro sistema ofrece a los trabajadores y sus dependientes una gama tan amplia de prestaciones básicas. La Comisión Nacional sobre la Seguridad Social ha calculado que la universalización de la cobertura incrementaría los ingresos del sistema en un promedio de 5000 millones de dólares anuales durante los próximos 75 años. Evidentemente, también aumentarían sus obligaciones.

Y por lo que respecta a los gastos, ¿existe algún procedimiento equitativo que permita reducir significativamente las futuras obligaciones del sistema? La respuesta es afirmativa. Existe un creciente consenso entre las personas que han estudiado el sistema en los últimos años en el sentido de que debería aplazarse la edad normal de jubilación de los 65 a los 68 años. El cambio sería doblemente favorable por cuanto aumentaría la población de cotizantes al mismo tiempo que reduciría la población de beneficiarios. Esta medida debería anunciarse con la debida antelación y sólo debería introducirse gradualmente, quizá con un aplazamiento de sólo algunos meses cada año. La Comisión Nacional calcula que este cambio representaría un ahorro de unos 11.000 millones anuales entre 1980 y 2055.

Actualmente el sistema tiende a favorecer la jubilación anticipada. Un trabajador puede jubilarse a los 62 años y percibir sólo un 20 por ciento menos cada año que si se hubiera jubilado a los 65. En consecuencia, unos tres de cada cuatro trabajadores se jubilan antes de cumplir los 65 años. El presidente Reagan propuso aumentar a un 45 por ciento la reducción de la pensión por jubilación anticipada, pero retiró la propuesta ante la enérgica oposición de la opinión pública.



RELACION entre beneficiarios y trabajadores, proporción clave que determina la solidez financiera del sistema. La relación ha aumentado de dos beneficiaros por cada 100 trabajadores en 1945 a 31 en la actualidad. Se da una proyección de los futuros incrementos, que implican un creciente desequilibrio entre pensiones e ingresos, hasta el año 2055 en base a supuestos intermedios (*negro*) y pesimistas (*color*).

Paralelamente al aumento de las jubilaciones anticipadas se observa un descenso de la disposición a trabajar pasados los 65 años. Mientras en 1960 un hombre de cada tres y una mujer de cada 10 continuaban trabajando después de la edad de jubilación legalmente establecida, actualmente sólo lo hacen un hombre de cada cinco y una mujer de cada 12. Esta disminución de la participación en la fuerza de trabajo, combinada con una mayor longevidad, genera una presión sobre el sistema. En 1940 sólo 583 hombres de cada 1000 y 687 mujeres entre 1000 llegaban a los 65 años; esas proporciones han aumentado hasta 711 y 839 por 1000, respectivamente. Menos de una tercera parte de todas las personas de 65 años o más tenían más de 75 años en 1974; hacia finales de siglo esta proporción será de casi la mitad.

Como ya he indicado antes, cualquier aplazamiento de la edad de jubilación tendría que introducirse gradualmente. Rita R. Campbell, de la Institución Hoover de la Universidad de Stanford, ha sugerido que se concedan condiciones especiales a las personas que han estado empleadas sin interrupción durante 45 años o más; señala que muchos planes privados de pensiones tienen en cuenta el número de años trabajados, además de la edad, al determinar la fecha de la jubilación. Muchas personas sufren problemas de salud que desaconsejan la prolongación de la actividad laboral pasados los sesenta años, sobre todo si se trata de un trabajo corporal pesado; tal vez podría modificarse la definición de invalidez para incluir tales casos.

Una alternativa al aplazamiento de la edad de jubilación (o quizás una medida simultánea) podría ser la adopción de una reciente innovación sueca llamada "jubilación por etapas". Las personas mayores pueden optar por realizar una actividad menos intensa o gradualmente decreciente en vez de pasar de la noche a la mañana del empleo a jornada completa a la jubilación total.

¿Podría ponerse coto al aumento de las obligaciones de la Seguridad Social haciendo frente a parte de las mismas con fuentes alternativas de ingresos para las personas jubiladas? La Comisión presidencial sobre Política de Pensiones recomendó el año pasado la implantación de un sistema obligatorio de pensiones privadas. Los empresarios pagarían una cantidad equivalente al 3

por ciento de su nómina de salarios; se introducirían cambios en la política fiscal destinados a desgravar a las pequeñas empresas. La comisión también insistió en la necesidad de modificar la estructura impositiva con medidas capaces de estimular el ahorro personal.

Es muy poco probable que el Congreso apruebe un plan obligatorio de pensiones privadas. Para empezar, existe una permanente desconfianza hacia los planes privados existentes; aunque el Congreso intentó normalizar estos planes y ampliar su cobertura en 1974, muchos trabajadores se quedan todavía sin cobrar nada o con una pensión muy reducida. Los planes privados de pensiones ya representan una pérdida de 15.000 millones de dólares anuales en ingresos fiscales para la nación. Los pequeños empresarios, muchos de los cuales operan casi en el límite de rentabilidad, se opondrían duramente a un plan obligatorio. Establecer un control social eficaz sobre los cuantiosos fondos de pensiones privadas ya constituye un problema y un nuevo programa lo exacerbaría aún más. Además, ¿para qué crear un nuevo sistema obligatorio cuando se puede lograr que funcione el actual sistema obligatorio, esto es, la Seguridad Social?

Muchos economistas alegan que en los Estados Unidos existe un nivel demasiado bajo de ahorros personales y que ésta es en gran parte la causa de una insuficiencia de capital, a la que será preciso poner remedio si se espera que el país modernice sus instalaciones industriales y maquinaria. (Martin S. Feldstein, de la Universidad de Harvard, se propuso demostrar hace algunos años que el sistema de la Seguridad Social era directamente responsable del bajo nivel de ahorro, pero muchos economistas piensan que se equivoca.) Independientemente de que se necesite o no un mayor ahorro para capitalizar adecuadamente la industria (y algunos economistas, entre ellos Thomas Juster, de la Universidad de Michigan, rechazan esta suposición), el Congreso ya ha ofrecido ventajas fiscales a las personas que ahorran en vistas a su jubilación: el Plan Keogh y la cuenta individual de jubilación (*Individual Retirement Account,* IRA) son dos ejemplos de ello.

Sin duda se ofrecerán en el futuro nuevos incentivos fiscales destinados a inducir a un mayor número de personas a ahorrar en vistas a su jubilación, pero sólo las personas con rentas elevadas pueden aprovechar plenamente tales oportunidades. Cualquier esfuerzo serio para inducir a los grupos con rentas medias y bajas a ahorrar una parte mayor de sus ingresos requerirá seguramente una subvención gubernamental y la protección de los pequeños ahorros contra la inflación, tal como se hace en Alemania. En todo caso, excepto para las personas muy pobres, las pensiones privadas y ahorros personales ya representan entre una cuarta y una tercera parte de los ingresos de las personas de edad. Es poco probable que consiga aumentarse esta proporción en más de algunas centésimas. La única alternativa es, por tanto, asegurar la solidez financiera del sistema de la Seguridad Social, al objeto de que éste pueda proporcionar unos ingresos básicos a las personas de edad.

Para lograrlo, en mi opinión, es preciso adoptar las siguientes medidas. Debería ampliarse el sistema de la Seguridad Social con la mayor rapidez posible hasta ofrecer una cobertura universal, lo cual incrementaría significativamente sus ingresos y ampliaría también la gama de prestaciones para las personas no incluidas todavía en el sistema. Debería tomarse pronto una decisión en el sentido de aplazar gradualmente la edad de jubilación, probablemente hasta los 68 años. La mitad de las pensiones que se sufraga a partir de rentas sobre las que no se ha percibido nunca ningún impuesto debería someterse al pago del impuesto sobre la renta, pero sólo en el caso de aquellas personas de edad cuyos ingresos totales sean lo bastante elevados para permitirles sufragar el impuesto. Hay que determinar los cambios a introducir en el seguro médico Medicare para que ofrezca un nivel aceptable de cuidados médicos a las personas de edad y actúe con mayor eficacia que hasta la fecha. Deberían empezar a rectificarse las injustas discriminaciones que sufren las mujeres en su condición de trabajadoras y de esposas. Convendrá reajustar nuevamente las pensiones de los trabajadores con largos años de servicio y un salario bajo, reajustadas ya hace poco, a fin de situarlas por encima del límite de pobreza.

El sistema de la Seguridad Social puede ofrecer medios básicos de subsistencia en la vejez a todos los estadounidenses, pero sólo a condición de que no se le imponga una carga excesiva. Por eso, la clave de la reforma está en estimular a la gente a trabajar durante más tiempo, con lo cual se asegurarán unos ingresos que les permitirán vivir cómodamente cuando se jubilen. Este es un objetivo razonable para la nación, y no sólo por razones económicas. Como dice el Talmud: Cuando dejas de trabajar, mueres.

El sistema español de Seguridad Social

La situación crítica que atraviesa la Seguridad Social española alienta un clima reformista en el que caben diversas interpretaciones. Los datos para la comprensión de la crisis no son sólo de tipo económico

Mercedes Alcover Ibáñez y Luis Vila López

El español que en 1982 acuda a una oficina bancaria para el cobro de su pensión, el que espere pacientemente su turno en la consulta de un ambulatorio y quien observe críticamente estos acontecimientos diarios, antes de dar su opinión sobre eso debe pensar un momento, mientras mira su reloj o acaricia un billete de mil pesetas en su bolsillo. En un minuto, la Seguridad Social española está gastando casi cuatro millones y medio de pesetas, de los que unos tres millones se dedican a prestaciones en metálico, mientras algo menos de un millón y medio –también por minuto–, está costando la asistencia sanitaria. Si para manejar cifras más familiares contempla su billete de mil pesetas, sabe que ese papel, en manos de la Seguridad Social, se va a cambiar en 644 pesetas para pensiones y 280 para asistencia sanitaria. De las 76 pesetas restantes, 12 se gastarán en servicios sociales, 34 en pagar la administración, 21 en nuevas inversiones y las 9 últimas en amortizaciones y otros.

¿De dónde salió ese dinero? De los cuatro millones y medio de gasto por minuto, los españoles que trabajan están poniendo tres millones setecientas mil pesetas. Del billete de 1000 pesetas que tiene en su mano la Seguridad Social, la fuerza laboral (trabajadores y empresarios) ha aportado 832 pesetas. Las 168 restantes proceden del estado y otros servicios.

No es demasiado ortodoxa esta manera de presentar el presupuesto de la Seguridad Social para 1982, ni tampoco la estructura poblacional española. Pero apuremos el ejemplo. El pensionista salió del Banco con una queja resignada y el paciente perdió la paciencia en la consulta. A uno y a otro habría que aumentar sus pensiones y atenderles mejor. Pero si la situación económica empeora, no hay trabajo para los que quisieran y el envejecimiento de la población lleva a engrosar las filas de pensionistas, el problema va a ser –lo es ya– muy grave. Decía el entonces ministro de Trabajo, Sanidad y Seguridad Social, Jesús Sancho Rof, al presentar el presupuesto del 82: "Si no se modifica el sistema, la Seguridad Social puede verse abocada a dedicar la totalidad de sus recursos a las clases pasivas, provocando un auténtico crack financiero". ¿De qué sistema hablaba el ministro? ¿Cuál fue su gestación y desarrollo?

Podríamos ver un primer paso en la Ley de Accidentes de trabajo de 1900. A la siniestralidad laboral se añaden otros riesgos, como las pensiones de retiro y la cuestión del aseguramiento del paro forzoso. En 1908 se crea el Instituto Nacional de Previsión. La iniciativa privada presenta sus Sociedades de Socorros Mutuos. Hay un largo camino que va desde el enfoque de voluntariedad y libertad hasta el concepto de seguro social obligatorio. En 1917 la Conferencia de Seguros Sociales concluye la necesidad de implantación gradual e independiente de cada seguro, en lugar de establecer un plan orgánico de seguros sociales. Así se llega a la II República en cuya constitución se precisan los ámbitos de la futura legislación social: seguro de enfermedad, accidente, paro forzoso, vejez, invalidez, muerte, protección a la maternidad, salario mínimo y familiar. Pero este programa no llegó a cumplirse y la acción normativa se centró en el accidente de trabajo y el paro forzoso.

Después, la guerra civil de la que sale "un nuevo estado" con mentalidad propia para definir el concepto y alcance de su previsión social, que en ningún momento desea parecidos con la anterior etapa. Desde el Fuero del Trabajo hay una enumeración de seguros sociales que se implantarán asistemáticamente al no ser su "seguro total" más que un deseo. Se perfeccionan diversas medidas de protección: subsidio familiar (1938), subsidio de vejez (1939), seguro de enfermedad (1943).

Paralelamente a la prestación de estos seguros sociales cobra fuerza una antigua figura asistencial, el mutualismo. Más cercano a cada colectivo, de carácter obligatorio, vinculado a ramas de industria o servicios, tiende a la totalidad en cuanto a los riesgos protegidos, centrándose básicamente en las pensiones de jubilación. En 1946 se crea el servicio de Mutualidades y Montepíos laborales, con el fin de sistematizar el mutualismo fomentando las relaciones de solidaridad entre las distintas mutualidades; en 1954 se aprueba el Reglamento General del Mutualismo Laboral. Otra creación de esta etapa es el Plus familiar (1945) que, financiado y organizado por cada empresa en particular, era un camino para llegar al salario familiar, y así conseguir "liberar a la mujer casada del taller y de la fábrica". En el Fuero de los Españoles (1945), se formula ya una relación más coherente de los seguros sociales y se abandona la idea de seguro total, dejando abiertos los posibles riesgos objeto de seguro social. La Ley de Principios del Movimiento Nacional (1958) plantea el derecho de todos los españoles a los beneficios de la Seguridad Social. La necesidad de unificar lleva el encargo al Instituto Nacional de Previsión (INP), en 1957, de elaborar un Plan Nacional de la Seguridad Social. Por falta de acuerdo, se elaboró en realidad un proyecto de ley de bases que culmina en la Ley de Bases de la Seguridad Social de 1963. Y así llegamos hasta hoy.

La preocupación por modificar la ac-

tual situación de la Seguridad Social es compartida universalmente, variando sólo el alcance de esa modificación. (Una Seguridad Social con posibilidades de éxito no es obra de un gobierno, ni de la patronal, ni de las organizaciones obreras. Es obra de todos, porque requiere acuerdos, pero también identificación popular.) Ante la falta de acuerdo del gobierno, las organizaciones empresariales y las centrales sindicales (febrero 1982), sería pretencioso mostrarnos en posesión de una piedra filosofal que los demás no han encontrado. Pero vamos a razonar para una mejor comprensión de la crisis actual a partir de: 1.º) la peculiar organización y objetivos de la Seguridad Social; 2.º) la composición de la población española; 3.º) la limitación de recursos que presentan las cotizaciones sobre los salarios y la aportación estatal; 4.º) la sobrecarga que suponen algunos capítulos del gasto de la Seguridad Social. Bien entendido que se trata sólo de algunos datos significativos para comprender la hondura de la crisis y su posible desenlace.

Vayamos con el primer apartado: organización y objetivos de la Seguridad Social.

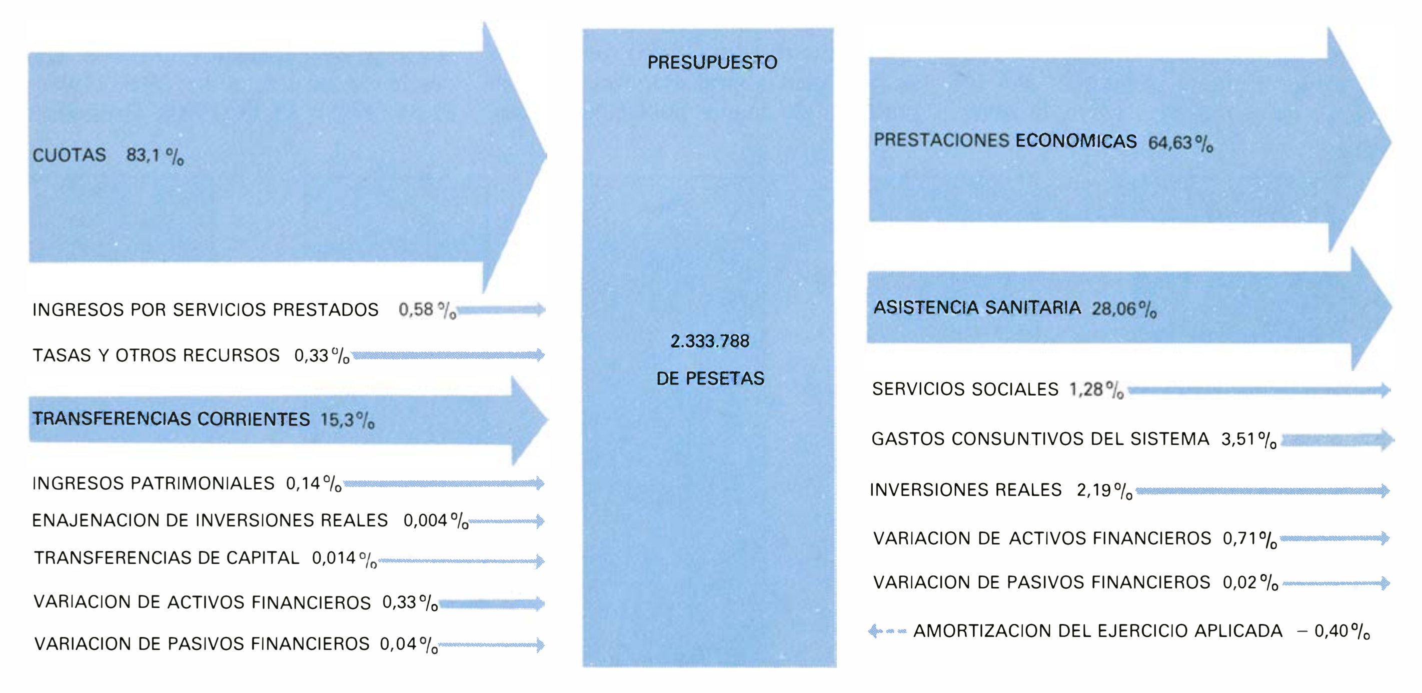
La organización del sistema es la resultante de su peculiar proceso de crecimiento y desarrollo. La característica central del sistema español de Seguridad Social es su fraccionamiento. Las sucesivas reformas introducidas, desde la Ley de Bases de 1963, al no replantear desde la raíz el tema de la estructura de la Seguridad Social, no han hecho sino aumentar la complejidad organizativa. ¿Cómo se halla estructurado? De acuerdo con un criterio rigurosamente profesional (según el tipo de actividad productiva), en el que se yuxtaponen un Régimen General, que ocupa una posición central y se configura como modelo en cuanto a ideal de cobertura del sistema, y unos Regímenes Especiales enunciados de forma abierta, es decir, susceptibles de ampliación.

El Régimen General agrupa a los trabajadores por cuenta ajena de la industria y los servicios, no expresamente incluidos en algunos de los Regímenes Especiales, y a ciertas personas asimiladas. Engloba también ciertos grupos profesionales de dudosa calificación laboral, a la vez que actúa transitoriamente como centro de organización de la protección de colectivos (clero, por ejemplo), respecto de los cuales se ha previsto la integración en un régimen especial no regulado todavía.

En los Regímenes Especiales que prevé la ley de Seguridad Social se distinguen cinco grupos: (*a*) regímenes de trabajadores por cuenta ajena: minería del carbón, ferroviarios, artistas, toreros, empleados del hogar, representantes de comercio, personal civil no funcionario al servicio de establecimientos militares y futbolistas profesionales; (*b*) regímenes de trabajadores por cuenta propia: autónomos y escritores de libros; (*c*) regímenes mixtos de trabajadores por cuenta propia y trabajadores por cuenta ajena: régimen especial agrario y régimen especial de trabajadores del mar; (*d*) regímenes de funcionarios públicos; y (*e*) regímenes de personas sin condición jurídica de trabajador o funcionario: estudiantes. Desde el punto de vista de la protección dispensada por la Seguridad Social, las diferencias se establecen no sólo atendiendo a la división entre trabajadores por cuenta propia y trabajadores por cuenta ajena, sino también entre los incluidos dentro de uno u otro de estos dos grandes grupos. Ocurre así que, en ocasiones, la asistencia sanitaria sólo se refiere a pensionistas (caso de representantes de comercio) o a pensionistas con ayudas económicas con motivo de intervenciones quirúrgicas para el resto (autónomos); la protección familiar, en el caso de los representantes de comercio, los artistas y los toreros, centrándonos en los trabajadores asalariados, se refiere sólo a prestaciones de pago único.

El déficit de los distintos Regímenes Especiales supone 560.000 millones de pesetas, que cubre en parte el superávit del Régimen General. La incapacidad para hacer frente a sus gastos con recursos normales y los desequilibrios demográficos conducen a aconsejar las diferentes fórmulas para la desaparición de esos Regímenes Especiales, que es a lo que se tiende.

La organización de la Seguridad Social tiene mucho que ver con los objetivos de la misma y su evolución.



RECURSOS Y GASTO del anteproyecto del presupuesto de la Seguridad Social española para el año 1982. En el primer sector de la ilustración puede apreciarse cómo el capítulo de las cotizaciones de empresarios y trabajadores alcanza el 83 por ciento. Frente a la opinión extendida de que el presupuesto se destina mayoritariamente a asistencia sanitaria, las cifras nos muestran que las prestaciones económicas consumen más de la mitad de los recursos.

Pensemos, además, que el volumen que ha adquirido la Seguridad Social en los países industrializados obliga a atender tanto a la protección social como a su incidencia en la economía general de esos países, integrando así política social y política económica.

Entre sus fines propios y directos están la administración de fondos para mejorar o consolidar la protección social lograda y garantizarla en el futuro, reducir las desigualdades que en la protección se han ido generando, repartir mejor la carga financiera y extender progresivamente sus prestaciones a toda la comunidad nacional. Los fines secundarios e indirectos están en estrecha relación con la situación de crisis de las economías occidentales: hay que promocionar el empleo o al menos seguir respecto a él un comportamiento neutral, servir como mecanismo redistributivo de la renta a través de sus prestaciones en metálico y en especie, manifestar la solidaridad social en el reparto de las cargas y prestaciones, promover cauces de participación social en la gestión del bienestar, ser un factor antiinflacionista y contribuir al equilibrio económico del país. Todos estos –y otros– fines secundarios actúan como condiciones económicas, a las que hay que sumar las políticas y sociales, para el logro de la maximización del bienestar social.

Concluyendo: Una complejidad en la estructura del sistema, innecesaria, que hace difícil la gestión, que da cabida a niveles protectores muy diversos y que engendra desequilibrios en la financiación.

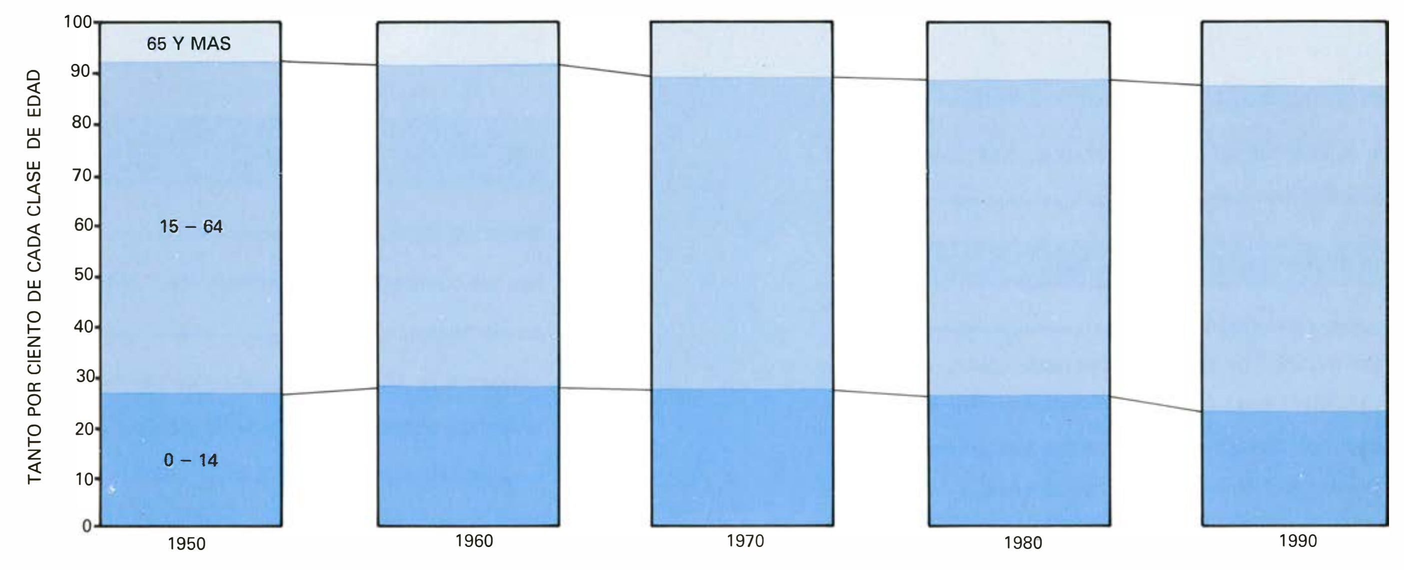
¿A qué población protegerá el sistema anteriormente descrito? De los datos demográficos seleccionamos, para su tratamiento, los siguientes: la población total española, la población activa y sus relaciones con la población total y la población protegida por la Seguridad Social. La razón de esa pormenorización radica en el efecto multiplicador que sobre la crisis presentan estas magnitudes demográficas.

La población española ofrece, en términos absolutos, unas perspectivas de crecimiento, aunque menos por el aumento de la natalidad que por el decremento de mortalidad: crece, pero con más viejos. La evolución de la estructura por edades, en porcentajes de la población total y para los tres intervalos confeccionados es suficientemente ilustrativa [*véase la ilustración de esta misma página*]. Hasta los catorce años se mantienen porcentajes similares, decrecen algo los correspondientes al intervalo 15-65 años (edad para engrosar la fuerza de trabajo), aumentan su proporción los mayores de 65, dentro de los cortes cronológicos seleccionados (adviertánse los índices de crecimiento de los intervalos de edad, base 1950). La prospección hasta 1990 agudiza las tendencias [*véase la figura citada*]: la población joven disminuye, el crecimiento de los mayores es notorio, mientras que el aumento de la población potencialmente activa (15-65) no es especialmente significativo.

Una nueva lectura de estas tendencias, con la vista puesta en las implicaciones económicas de tales estructuras de edades, lleva a señalar nuevos aspectos. Por una parte, la posibilidad del paro que prende, lógicamente, en el intervalo intermedio (15-65), reduciendo su capacidad laboral. Por otra, la incorporación de nueva población pasiva, superior al volumen de la potencial incorporación al mercado de trabajo que representa la promesa biológica del primer intervalo. Además, el aumento de la esperanza de vida al nacer [*véase la ilustración superior de la página 20*] que es propia de los países desarrollados y que, en treinta años de diferencia (1960-1990), se sitúa en 73, 79 años. Por último, quien se haya jubilado en 1975, puede pensar que todavía le quedan quince años de vida [*véase esa misma ilustración*].

La estructura por edades de la población española no solamente permite prever unas necesidades futuras a las que el sistema tendrá que hacer frente. Dirigir la mirada hacia algunos aspectos de la población activa (sin perder de vista las cifras de paro registrado que rebajan el número de población activa real) y relacionarlos con la población total y, dentro de ella, el intervalo que representan los no-activos (menores de 15 y mayores de 65 años) nos presenta el problema de las tasas de dependencia. En otros términos, la capacidad para generar recursos vía cotizaciones o cualquier otra alternativa impositiva fiscal.

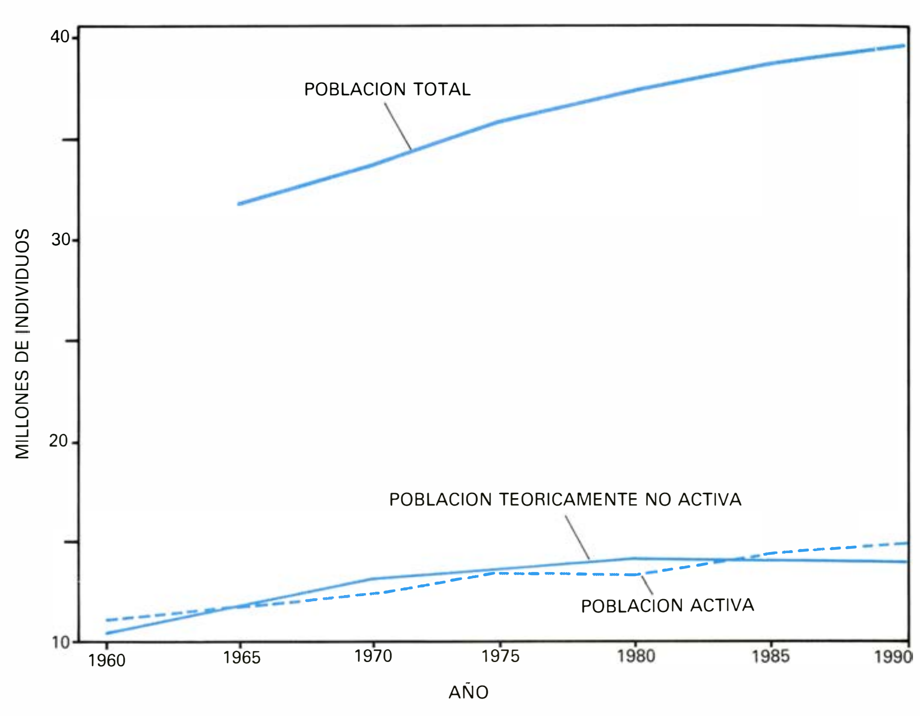
De cada cien españoles, alrededor de 37 eran potencialmente activos entre 1965 y 1975 [*véase la ilustración de la derecha*]. El año de 1980 marca un año de crisis y las estimaciones para 1985 y 1990 prevén la vuelta a las cotas de 1965. El panorama puede ser más oscuro, si lo hacemos en términos de población ocupada. Si se mantiene la cifra de paro estimada para junio de 1981 en 1.783.500 parados, según el Instituto Nacional de Estadística, la proporción desciende sensiblemente: 29,93 (1980); 32,54 (1985); 33,33 (1990). Estabilizar



LA POBLACION ESPAÑOLA ofrece, en términos absolutos, una perspectiva de crecimiento, aunque menos por el aumento de la natalidad que por el decremento de mortalidad. Hasta los catorce años se mantienen porcentajes similares, decrecen algo los correspondientes al intervalo 15-65 años (posible edad para engrosar la fuerza de trabajo) y aumentan su proporción las personas mayores de 65 años. La prospección hasta 1990 agudiza las tendencias.

la situación de paro es una hipótesis optimista. Según el ministerio de Economía, en seis años (1975-1981) la cifra de parados se ha triplicado: 1975, 510.500 parados (base 100); 1980, 1.620.300 (índice 317,39); 1981 (primer trimestre), 1.741.300 (índice, 341,10). La relación entre población ocupada y población total (en otros términos, población potencialmente cotizante a la Seguridad Social y población potencialmente beneficiaria de las prestaciones de la Seguridad Social) es cada vez más pequeña, con el consiguiente problema financiero de acopio de recursos.

Una última reflexión sobre la población activa y la población no activa (menores de 15 años y mayores de 65) [*ilustración de la derecha*]. La población no activa es teóricamente dependiente. Y esta relación de dependencia, considerando los ritmos de crecimiento de ambas magnitudes, pasa en 1980 por un punto crítico, especialmente si hacemos la misma reflexión sobre la incidencia del paro en las cifras de población activa.



EVOLUCION DE LA POBLACION total, activa y teóricamente no activa. El ritmo de crecimiento de la población activa es inferior al de la afiliación, como consecuencia de la generalización progresiva de ésta al incluir colectivos nuevos. La población no activa, agravada con el paro, es teóricamente dependiente.

De todas las anteriores consideraciones sobre la incidencia social y económica de las características de la población española y su aspecto inductor, e incluso multiplicador, de la crisis, se llega a la formulación de la "tasa de actividad total" [*véase la ilustración inferior de la página siguiente*], que muestra una tendencia a la disminución para los varones y para el total; la tendencia negativa no afecta al trabajo de la mujer, y no obstante el incremento de su tasa, éste es insuficiente para compensar la progresiva disminución de la tasa masculina.

El cumplimiento de la finalidad protectora de la Seguridad Social se muestra a través de la comparación entre población total y población protegida de hecho; también, dado el carácter profesional del sistema español de Seguridad Social, mediante comparaciones entre el colectivo de afiliados (sin sus beneficiarios) y el volumen de población activa que marca el techo máximo de los niveles posibles de afiliación. Empezamos por lo segundo [*véase la ilustración de la página 21*].

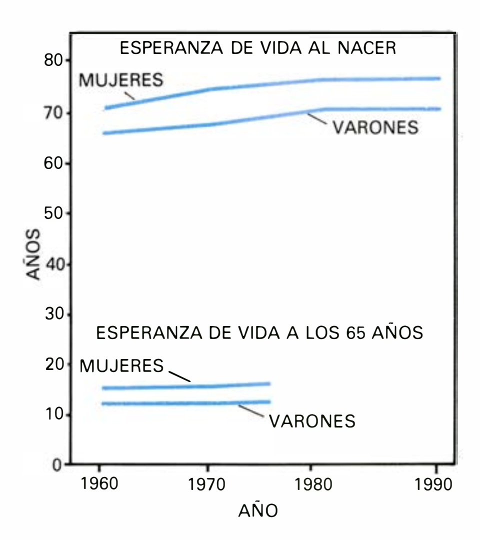
Primeramente se alude a "población activa" (no a población ocupada) y se contempla en el colectivo de afiliados tanto la posibilidad de la doble afiliación en distintos regímenes, como la carencia de afiliación en colectivos que deberían estar incorporados (por ejemplo, el servicio doméstico). Todo esto explica y equilibra una posible distorsión en ambos efectivos (activos y afiliados). Las tendencias se advierten de inmediato. El ritmo de crecimiento de la población activa (1970, base 100) es inferior al de la afiliación, como consecuencia de la progresiva generalización de ésta a través de la incorporación de colectivos antes no incluidos. Como resultado, el ritmo de crecimiento de la cobertura de la población activa por la población afiliada es favorable a la deseada ecuación entre los mismos. La aceleración mentada es más sensible entre 1970-1975 que en el quinquenio posterior a esta fecha.

Por último, la relación entre población total y población protegida (afiliados más sus beneficiarios) que definirá la pretensión universalista del sistema y la consecución definitiva de su fin protector. Si hay colectivos protegidos, aunque no lo sean en asistencia sanitaria, el número de afiliados y beneficiarios habría de incrementarse, con lo que aumentaría la cobertura sobre la población total. Utilizando esta hipótesis, nos atrevemos a afirmar una cobertura del 86-87 por ciento en 1980. La realización definitiva de la cobertura universal equivaldría a completar ese 14-13 por ciento restante. El índice de variación (1970, base 100) de la población protegida es mucho mayor que el de la población total. Luego la cobertura progresa satisfactoriamente. En los últimos cinco años (1975-1980) es todavía superior al del primer quinquenio.

¿De qué recursos dispone nuestra Seguridad Social para subvenir a la población descrita? De los recursos de la Seguridad Social, en 1982, el 83,19 por ciento (casi dos billones de pesetas) procederán de las cotizaciones de empresarios y trabajadores; el 14,99 serán subvenciones del estado (trescientos cincuenta mil millones); el restante 1,82 por ciento provendrá de otros recursos.

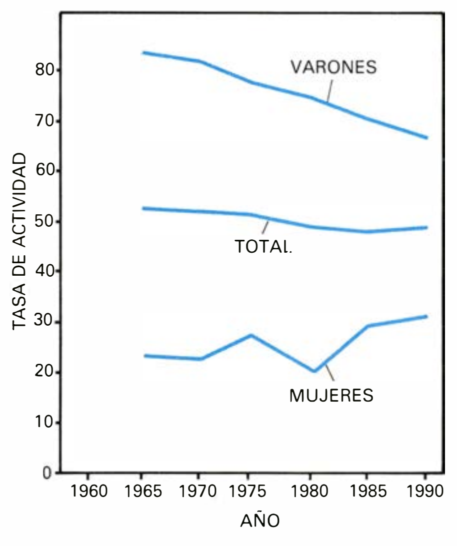
De los casi dos billones que gravan el factor trabajo, 6.989.787 cotizantes al Régimen General aportan el 76,41 %, mientras que 3.730.267 cotizantes a los distintos regímenes especiales lo harán en un 15,44 por ciento. Por cada 10 afiliados a la Seguridad Social, seis y medio lo están en el Régimen General y tres y medio en los especiales. Pues bien, de 1000 ptas. que por cotizaciones ingrese la Seguridad Social, más de 750 las pondrán seis y medio trabajadores del Régimen General y apenas 150 los trabajadores de los Regímenes Especiales. La desproporción en cuanto capacidad contributiva es evidente. Y la tendencia arranca desde los comienzos de la actual configuración por regímenes. La unidad de caja de la Seguridad Social permite que el déficit (gastos/recursos) de los Regímenes Especiales se enjugue con los superavit del Régimen General, si es que existen. El peso, porcentual, que representan las cotizaciones no ha variado significativamente en los últimos años.

Esta enorme carga contributiva se gira sobre el factor trabajo, y no es extraño que las fuerzas sociales afectadas intenten una traslación de la misma, del



ESPERANZA DE VIDA al nacer y a los 65 años. Las curvas muestran unos valores comparables con los que presentan los países desarrollados.

trabajador a la empresa, de la empresa al estado. Ciertamente, esta forma tradicional para generar recursos presenta, junto a indudables ventajas recaudatorias (por ejemplo, la amplia base que supone el número de trabajadores y el peso de las remuneraciones al factor trabajo dentro de la renta nacional, el interés de los trabajadores en que se paguen puntualmente las cuotas, etc.), otros inconvenientes. Quizás el mayor es la discriminación contra el factor trabajo y, en concreto, contra las empresas en que la relación capital/trabajo es menor. Este encarecimiento del factor trabajo, no desgravable ante la exportación, disminuye la competitividad exterior de nuestros productos. Las economías transnacionales van a inclinarse por una política de industrialización de la periferia, buscando una mejor economía de costes. El empleador nacional, si tal medida no está a su alcance, lo pensará dos veces a la hora de sustituir la fuerza de trabajo en vísperas de su jubilación.



DE CADA CIEN ESPAÑOLES, alrededor de 37 eran potencialmente activos entre 1965 y 1975. El año de 1980 marca una fecha de crisis y las estimaciones para 1985 y 1990 prevén la vuelta a 1965.

El porcentaje que representan las cotizaciones a la Seguridad Social sobre el producto interior bruto (PIB) era casi el 11 por ciento en 1979, superior a la mayoría de los países de la OCDE (media 9,29), aunque algunos le sobrepasen. Pero esta comparación debe hacerse simultáneamente respecto a lo que suponen *todos* los ingresos fiscales (incluidas las cotizaciones a la Seguridad Social) respecto al PIB. En este caso, España ocupa el último lugar (España 22,81, media OCDE 31,14) en cuanto al peso de fiscalidad total que soporta (impuestos más cotizaciones de Seguridad Social).

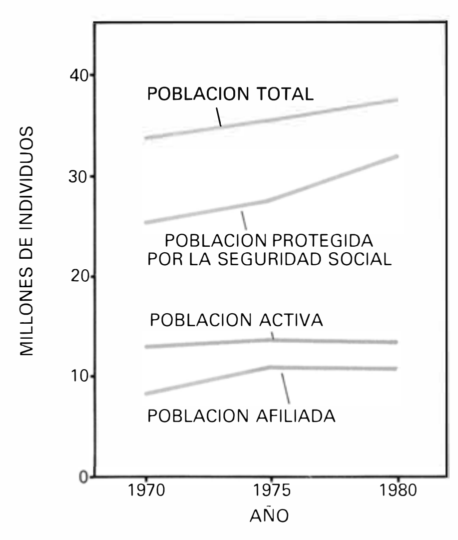
El problema de la insuficiencia o injusticia recaudatoria de las cotizaciones como fuente primordial de recursos para la Seguridad Social se completa con el análisis de la distribución que de tal carga se hace entre empresa y trabajador. Para el conjunto de las cotizaciones dentro del presupuesto, los trabajadores aportan el 22 por ciento y las empresas el 78 por ciento. Al considerar conjuntamente las cargas fiscales con las de Seguridad Social se advierte que la proporción en el pago es superior en España para empleados y empleadores, si bien las diferencias respecto a los valores de la CEE y la OCDE son superiores en el caso de los empleadores. La queja de éstos tiene motivos, aunque menos en los términos de cargas fiscales totales que de Seguridad Social.

El porcentaje que para 1978 representaban las cotizaciones de Seguridad Social sobre el total de cargas fiscales era mayor para España (trabajadores, 9,79 por ciento; empresas, 39,71 por ciento) que para la CEE (trabajadores, 8,75 por ciento; empresas 16,39 por ciento) y para la OCDE (trabajadores, 6,92 por ciento; empresas 16,12 por ciento). Si se paga más a la Seguridad Social, se paga menos al estado y son menos sus posibilidades financiadoras.

Ante esa situación, se proponen modificaciones posibles dentro del mantenimiento de la base de cotización en función del salario, vía aumento de los tipos de cotización sobre determinadas remuneraciones o elevación de los topes máximos de cotización. Dada la importancia de las cotizaciones como fuente recaudatoria, pensar en un tope mínimo exento provocaría un grave drenaje económico del sistema. Ante la tendencia sustitutoria de trabajo por capital por parte de las empresas, urge imaginar bases de cotización alternativas que graven mejor la capacidad de pago de la empresa sin penalizar la mano de obra. En busca de lo viable, y no sólo de lo teóricamente posible, se habla también de la utilización del futuro impuesto del valor añadido (IVA) como base para las cotizaciones. Esto por lo que hace referencia a los ingresos de la Seguridad Social a través de las cotizaciones. ¿Qué decir de la segunda fuente de financiación, el estado?

El incremento de las aportaciones estatales, vía presupuestos generales del estado, es una petición constante por parte de las fuerzas de trabajo, con el fin de aligerar sus cargas y homologarse con las participaciones estatales en los sistemas europeos. La convergencia existente entre los sistemas de corte anglosajón y el deseo de las empresas apunta en este sentido. La tendencia a que los beneficios de la Seguridad Social lleguen a todos, también. Esta aportación estatal puede discurrir por una doble vertiente: la primera, asumiendo el estado por sí mismo determinadas prestaciones que anteriormente eran responsabilidad de la Seguridad Social (por ejemplo, los subsidios al desempleo). La segunda, mediante transferencias corrientes a la Seguridad Social para el logro de sus cometidos. Pero, en cualquiera de los dos casos, piénsese que la Seguridad Social no es la única expresión de la polítia social de un estado. Fuera del presupuesto de la Seguridad Social, otras instituciones públicas y privadas otorgan prestaciones que, junto con las de la Seguridad Social, configuran el llamado presupuesto social de un país. El presupuesto social europeo ya existe: la estimación del español está en marcha. Lo que la España de 1980 gastó en prestaciones sociales rondaba el 16,8 por ciendo del PIB. El 70 por ciento de ese gasto correspondía a la Seguridad Social. Los países de la CEE dedican a estos gastos entre el 20 y el 30 por ciento de su PIB. (El peso que el presupuesto de Seguridad Social ha ido adquiriendo respecto a otras magnitudes económicas queda reflejado en la ilustración inferior de la página 22.)

La Seguridad Social española, como todas las del área comunitaria, está fundada sobre un sistema de reparto; de lo que *hoy* unos pagan, *hoy* se benefician otros; el día que yo reciba prestaciones, otros estarán pagando por mí. En momentos de crecimiento económico, de auge sostenido, de pleno empleo, de bajas prestaciones, el volumen de ahorro generado puede acudir a alimentar las inversiones del país. Pero cuando ya

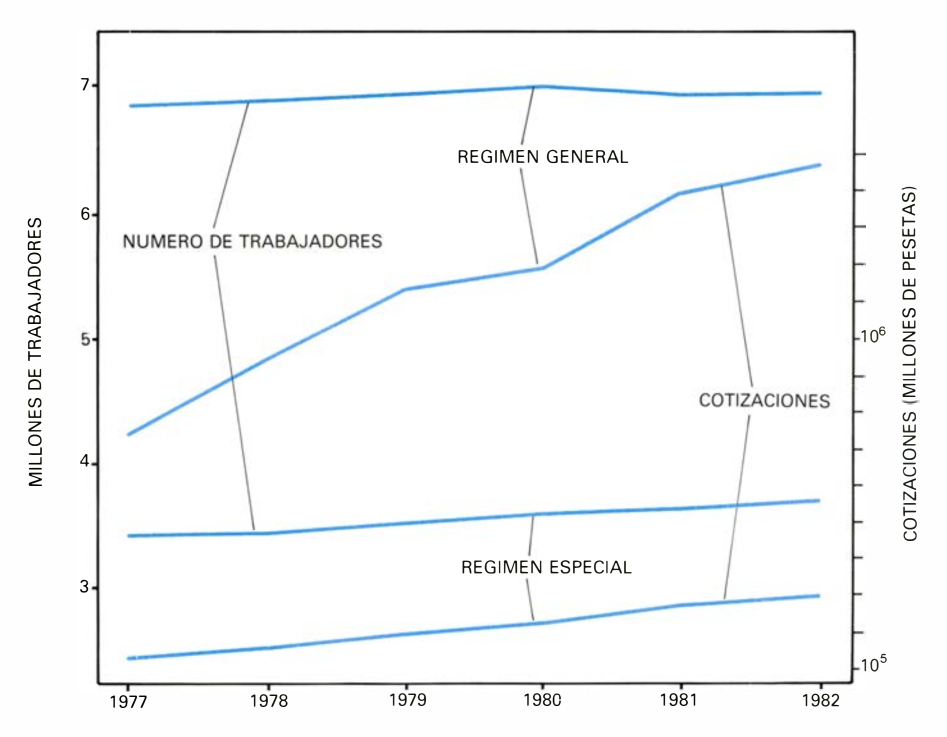


CRECE más deprisa la extensión de la cobertura social que la población total, ritmo que se evidencia con especial firmeza en el último quinquenio.

no es ésta la situación, la solvencia económica del sistema está en peligro. La Seguridad Social creció en los años de auge; los problemas aparecen ahora con la crisis económica generalizada.

La otra fórmula, hoy casi desaparecida y no recomendada por la OIT, es un sistema de capitalización que garantice en cualquier momento las obligaciones contraídas. La pérdida de valor de sus activos es el problema. Bajo cualquiera de las dos fórmulas, o las posibilidades combinatorias intermedias, hay que buscar una garantía futura de recursos. La capacidad de ahorro de la Seguridad Social española ha sido menor que las inversiones necesarias para su funcionamiento, por lo que ha tenido necesidad de financiación externa desde 1976 a 1979. La asunción de la ayuda al desempleo por parte de los presupuestos del estado alivió la carga posteriormente, pero lo que se ha hecho no es sino trasladar el déficit contable de la Seguridad Social a otras instituciones estatales.

¿Hasta dónde puede llegar el déficit por el gasto público en España? No es sólo la Seguridad Social la que espera una aportación del estado. A ella hay que sumar los resultados económicos de los pactos políticos de toda índole que se han concertado, que evidentemente tienen un coste. Por ejemplo, el Acuerdo Nacional de Empleo (ANE), la reconversión industrial o los Fondos de Compensación Interterritoriales derivados de la construcción del estado autonómico. El déficit previsto en los prespuestos generales del estado para 1982 se sitúa en 697.000 millones de pesetas (hay quienes lo estiman cercano al billón) financiados por deuda pública interior y exterior (227.000 millones), deuda del Tesoso (120.000 millones



DE LOS CASI DOS BILLONES que cotizan trabajadores y empresarios, los adscritos al Régimen General cargan con el 76,41 por ciento, y sólo un 15,44 % los cotizantes a los distintos Regímenes Especiales.

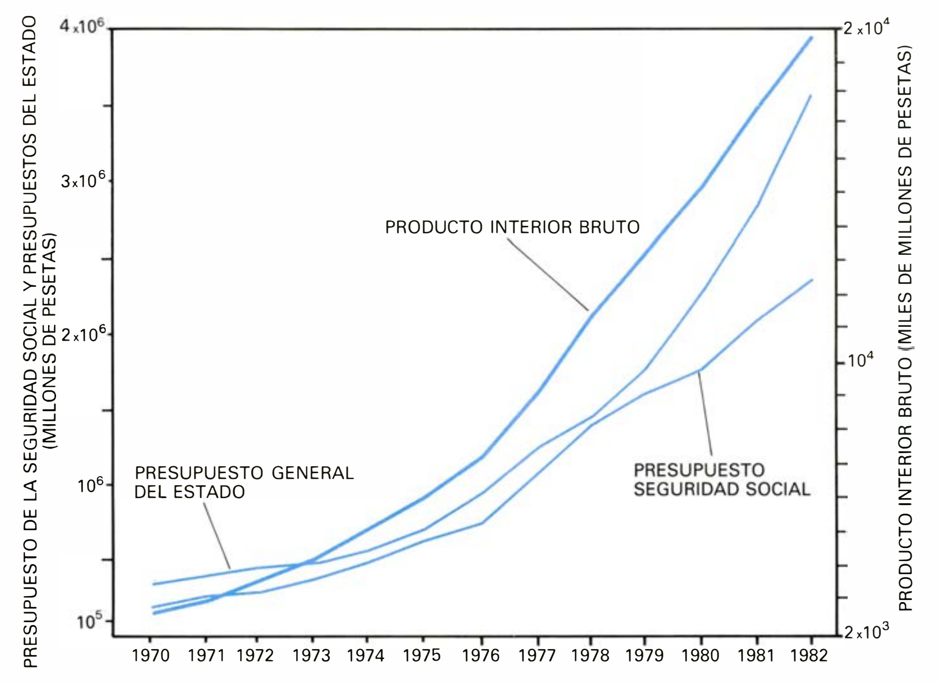
de pesetas) y apelación al Banco de España (350.000 millones). Los presupuestos son equilibrados, pero el impacto real de los ajustes contables sitúa ante una inflación que no perdona.

Recapitulando: Si las cotizaciones como fuente fundamental de recursos están cerca de su techo, si las prestaciones alcanzadas difícilmente admiten un recorte significativo, la única salida estará en incrementar las aportaciones estatales. Dejando aquí de lado el análisis de las prestaciones, el dilema se plantea como una exigencia de racionalidad en la selección del gasto público. Esta selección depende del programa social de la fuerza que ostente el poder.

En España, limitarse a enunciar la necesidad de una mayor aportación estatal es un simple traslado del problema, conocida la escasa capacidad fiscal del estado para todas las demandas que recibe. El incremento de recaudación de la Reforma Fiscal de Fernandez Ordoñez significó, como ejemplo, que siguiendo los acuerdos del Programa Económico del Gobierno (PEG), en 1979, el aumento en las transferencias a la Seguridad Social supuso los dos tercios del aumento de presión fiscal de los dos últimos años. La necesidad de acudir a nuevos impuestos en la línea de los indirectos parece confirmarse.

Tras los recursos, el gasto. La racionalidad del gasto público acompañada por la racionalidad mayor en la Seguridad Social no se detiene en temas de organización y estructura. La clasificación del gasto de la Seguridad Social, por sus funciones, incluye las prestaciones económicas, la asistencia sanitaria, los servicios sociales, los gastos consuntivos del sistema, las transferencias externas, las reservas, las inversiones reales y la variación en activos y pasivos financieros. Las dos primeras funciones (prestaciones económicas y asistencia sanitaria) se llevan el 92,31 por ciento del presupuesto y por ello merecen un desglose. El español medio piensa que los gastos en asistencia sanitaria absorben casi todo el presupuesto de la Seguridad Social y además aumentan vertiginosamente, cuando la realidad nos lleva a la reflexión sobre las prestaciones económicas. En éstas se incluyen las pensiones, la incapacidad laboral transitoria, la invalidez provisional, la protección a la familia, las indemnizaciones y entregas únicas reglamentarias y los auxilios económicos; pero son las pensiones las que se llevan la mayor parte del presupuesto para prestaciones económicas en 1982 (85,12 por ciento). Esta tendencia se acentuará en el futuro debido al progresivo envejecimiento de la población. En el término pensiones se incluye la jubilación (53,9 por ciento), invalidez (26,7 por ciento), viudedad (16,8 por ciento), orfandad (2,2 por ciento) y favor familiares (0,4 por ciento).

¿Qué perspectivas de evolución se divisan? La primera, el indudable incremento que deberá sufrir el volumen total dedicado a pensiones, debido a la natural elevación de las mismas, al envejecimiento de la población y a la generalización de este tipo de prestación a situaciones de desprotección tanto en algunos regímenes como en colectivos no integrados en el sistema. La segunda, incremento de beneficiarios: la progresiva esperanza de vida juega a favor de la prolongación de las percepciones de pensión; los deseos generalizados en amplias capas de opinión respecto a anticipación de las jubilaciones constituyen otro aspecto que hay que atender. La tercera, la pensión media, en cuyo carácter promedio se ocultan grandes desigualdades. La posibilidad



COMPARACION entre el presupuesto general del Estado, el PIB y el presupuesto de la Seguridad Social. El porcentaje que suponen las cotizaciones a la Seguridad Social sobre el PIB era, en 1979, de un 11 %.

de establecimiento del nivel fundamental y mínimo para toda la población constituye otro factor agravante en este capítulo protector.

Una discriminación contra los perceptores de prestaciones a largo plazo de la Seguridad Social aparece en la fórmula actual de revalorización de las pensiones, en oposición a la que de modo automático se establece para las rentas de los activos (art. 27 del Estatuto de los Trabajadores). Conocido el ritmo de la inflación y la capacidad de autodefensa de los pensionistas, cambiar el ajuste automático de las pensiones (en cuanto procedimiento, mecanismo y porcentaje de revalorización) por una adaptación "según las circunstancias" equivale a crear un disgusto potencial y larvado que, en una sociedad democrática, se paga caro en las elecciones. Y lo mismo vale, mutatis mutandis, en el problema más general de la protección a la vejez. Los temas de edad de jubilación, la compatibilidad entre trabajo y percepción de pensiones, tienen repercusión económica de relevancia.

Piénsese en los proyectos de jubilación anticipada y la propuesta estadounidense, explicada en otro artículo de este número, de signo absolutamente contrario: retrasar la edad de jubilación significa reducir la inflación, aliviar la tasa de dependencia pensionistas/activos, sanear la economía de la Seguridad Social, facilitar el pago gradual desde la situación de activo a pasivo, hacer posible el trabajo a tiempo parcial con reducción en las cotizaciones, y un largo etcétera. Aducir la posibilidad de "alivio del desempleo" al quedar libres puestos de trabajo, plantea dos preguntas: ¿cuántos puestos de trabajo se amortizan con la jubilación? y ¿qué cuesta más a la Seguridad Social, un jubilado o un parado? A título de muestra, la reducción de la jubilación en un año (64 años) costaría a la Seguridad Social española veinte mil millones más por año.

Aparte de la cuantía dedicada a las pensiones, otros dos epígrafes del gasto merecen un comentario: la asistencia sanitaria y los subsidios de desempleo.

Porcentualmente, dentro del presupuesto, ya vimos cómo las prestaciones económicas tenían un peso mayor (64,37 por ciento frente al 27,94 por ciento de la asistencia sanitaria, para 1982) sin restar importancia a los gastos en salud. Pero dada la importancia que el ciudadano concede a la protección de su salud, es inviable pensar en un recorte significativo en calidad y cantidad de

los cuidados que requiere. El volumen de medios técnicos y humanos que el cuidado de la salud exige invita, en primer lugar, a una mayor racionalización en el empleo y planificación de los mismos.

Además, el alivio del déficit financiero, vía recorte de los gastos sanitarios, casi nos parece técnicamente imposible. Por diversas razones, su ritmo de crecimiento puede ser superior al de las tasas de crecimiento económico del país. La razón intrínseca a la naturaleza de estos gastos podría ser de tipo tecnológico (equipamiento, formación) o bien demográfico (envejecimiento de la población, tipo de vida en las sociedades industriales, y otros). Volvemos, pues, a recurrir a la mejora interna del sistema de salud como camino para la reducción del gasto (ejemplo: terminar con los conflictos de competencias entre entes públicos que fomentan la dispersión y duplicidades innecesarias; terminar con el divorcio sanidad asistencial/sanidad preventiva; control mayor del gasto farmacéutico, etcétera). El proyecto de una asistencia sanitaria generalizada a toda la población y que se financie por vía fiscal permitirá una gestión tan independiente como la que posee la enseñanza y en la que pueden concurrir tanto la iniciativa privada como la gestión pública.

El Convenio 102 de la OIT trata la protección al desempleo como una de las típicas de la Seguridad Social y también la Constitución Española (art. 41) lo reconoce como derecho. Sin embargo, el alcance de la política de empleo, la necesidad de trasvases sustanciales (aparte de la recaudación obtenida por un tipo del 4 por ciento sobre la base hasta 31-XII-81) para la financiación del subsidio y los conflictos de racionalidad que tantas veces existen, adscriben la gestión de las prestaciones para el empleo al INEM (Instituto Nacional de Empleo), configurado como un organismo autónomo administrativo.

Hasta 1975, el paro registrado no constituye un problema especialmente grave: 301.852 parados que representaban el 2,25 por ciento de la población activa. ¿Y a partir de entonces? El ritmo crece hasta superar hoy la decena por ciento. Mas el paro subsidiado es muy inferior al paro real. Al haber configurado un modelo de "seguro social", éste resulta insuficiente ante la magnitud de un problema que no es algo coyuntural sino que responde a una crisis estructural del modelo económico. Fuera del subsidio quedan los que buscan su primer trabajo (nunca cotizaron a la Seguridad Social), pero también determinados grupos de profesionales que, en la variedad de prestaciones de los distintos regímenes, quedan excluidos; por ejemplo: empleados del hogar, trabajadores autónomos, representantes de comercio, etc. Así pues, una cobertura baja, y no tanto por la cuantía de las prestaciones cuanto por la tasa de cobertura del paro registrado por el seguro de desempleo. En el último año que este subsidio se incluía en las cuentas de la Seguridad Social (1979) la tasa aludida alcanzaba el 52,7 por ciento, lo que suponía, porcentualmente, sobre el total de prestaciones económicas y sanitarias, el 13 por ciento.

No puede cerrarse esta revisión de algunas variables de la Seguridad Social que están en la raíz de la crisis sin mentar aquella otra que está en la base de las posibles soluciones. Como la Seguridad Social es política social, ésta se articula con todas las políticas de un estado (monetaria, fiscal, internacional, etc.) La naturaleza de esta articulación no es un problema técnico, sino político. Ya no son cuestiones de producción, sino de distribución.

Así, la petición de una mayor aportación estatal disminuyendo la cuantía de las cotizaciones, la extensión a todos los ciudadanos y la adecuación de las prestaciones del sistema abocan ciertamente a un crack de la Seguridad Social, pero no tienen por qué llevar a una bancarrota social. Es el problema del reparto de los bienes escasos que da origen a la ciencia económica. Y el problema del reparto es siempre problema de poder. De esta manera nos encontramos, al término del artículo, ante una crisis económica del sistema de Seguridad Social que es en realidad una crisis contable del mismo. Revisar los asientos contables puede ser el momento para sugerir, por ejemplo, el problema político de los impagados a la Seguridad Social, el problema político de las clientelas más fuertes en la adjudicación del pastel presupuestario, el problema político de la selección de objetivos prioritarios. Se dirá que hay mucho "problema político". Desde Aristóteles, los autores están con los que creen que político es todo, porque social es la vida del hombre.

INCAPACIDAD LABORAL TRANSITORIA
OTRAS
8.77 %
6.11 %
PRESTACIONES SANITARIAS
23,7 %
PRESTACIONES ECONOMICAS
54.7 %
85.12 %
OTRAS
6.5 %
PENSIONES

PESO RELATIVO DE LAS DISTINTAS PRESTACIONES económicas, para el año 1982. Entre éstas se incluyen las pensiones, la incapacidad laboral transitoria, la invalidez provisional, la protección y otras.



Extinciones en masa del Mesozoico tardío

Muchas especies vegetales y animales, los dinosaurios en particular, desaparecieron de pronto hace 63 millones de años. Un estrato rico en iridio sugiere que la causa quizá fuera la caída de un asteroide

Dale A. Russell

Uno de los hechos más llamativos que nos revela el registro fósil de la vida de nuestro planeta es el de la desaparición simultánea de varias clases de reptiles, de invertebrados marinos y de plantas primitivas en las postrimerías de la era Mesozoica, hace de ello unos 63 millones de años. Durante generaciones enteras los paleontólogos han buscado sin éxito la solución del enigma. Pero tienen ahora en la mano varias pruebas que apoyan una hipótesis nueva: las desapariciones fueron el resultado de una perturbación de la biosfera por un agente extraterrestre.

El catastrofismo no constituye ninguna doctrina inédita en la historia de la descripción de los episodios por los que ha atravesado la Tierra. Pero no podemos decir de ella que haya gozado del aplauso de la comunidad científica. A principios del siglo XIX, cuando la geología daba sus primeros pasos, el anatomista francés Georges Cuvier propuso que el pasado había estado marcado por una serie de "revoluciones", o catástrofes, ambientales. Desde su punto de vista, tales perturbaciones habrían motivado tres desapariciones animales: la de los mamuts, al final de la edad del hielo, la de varios mamíferos primitivos, fosilizados en rocas muy por debajo de las gravas de la edad del hielo, y la de los reptiles gigantes, fosilizados en capas de creta situadas todavía a mayor profundidad. En las décadas inmediatamente siguientes, sin embargo, el trabajo de geólogos pioneros, como Charles Lyell, puso de manifiesto que los procesos de cambio de la historia de la Tierra se prolongaban mucho más de lo que Cuvier sostuviera. El catastrofismo decayó, ocupando la doctrina del gradualismo su lugar. Desde hace más de un siglo, los paleontólogos han aceptado, en general, que cualquiera que fuera la causa de las desapariciones de finales de la era Mesozoica, no pudo tratarse de un cataclismo mundial.

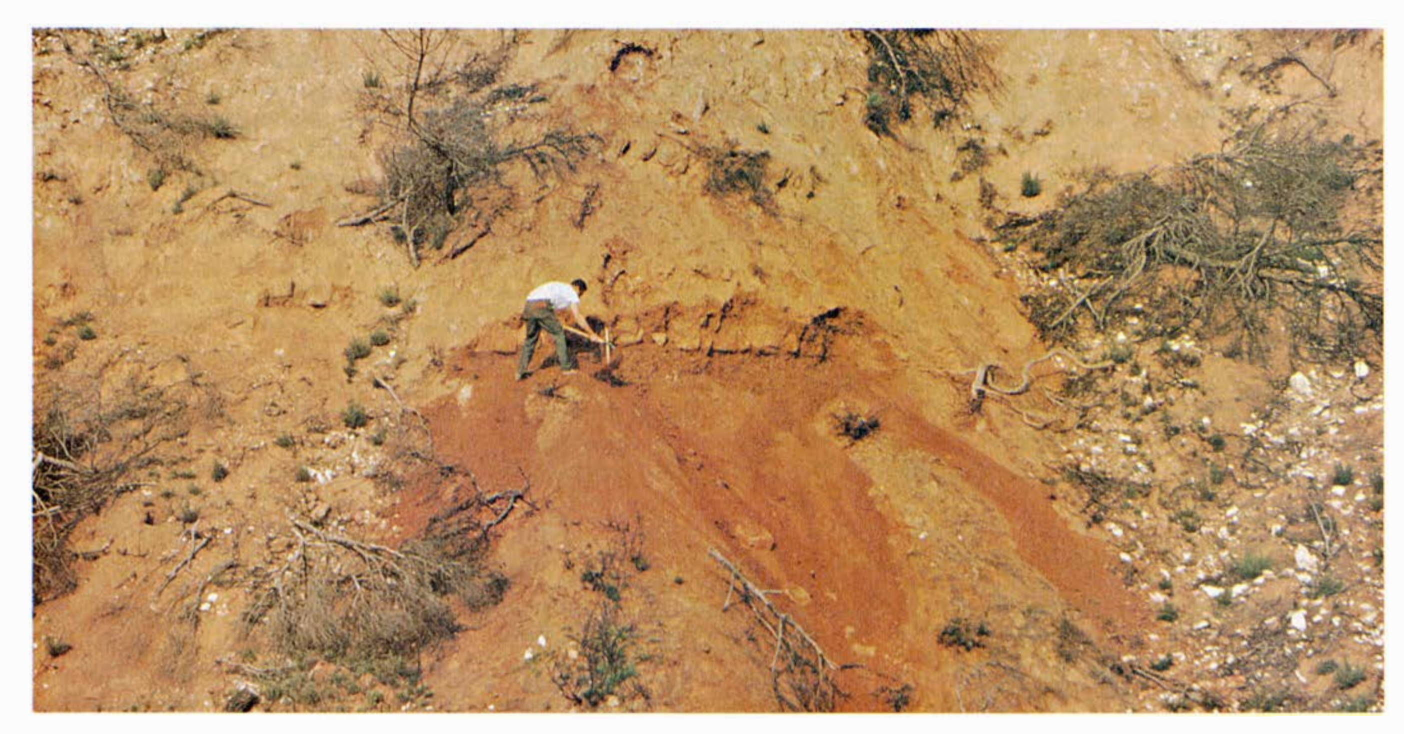
Las principales víctimas entre los reptiles fueron los dinosaurios. Valga de ejemplo saber que, en las postrimerías del período Cretácico, último capítulo del Mesozoico, por lo menos 15 familias no emparentadas de dinosaurios, que vendrían a constituir entre 50 y 70 especies distintas, poblaban América del Norte. En las rocas que se formaron inmediatamente después de ese período, no se ha encontrado esqueleto alguno de dinosaurio. Por cuya razón se identifica el final del Mesozoico con el momento en que se extinguieron los dinosaurios. Mas no se crea que fueron éstos los únicos organismos en desaparecer. Entre las otras 33 familias de reptiles que vivían en América del Norte a finales del Cretácico hubo que contar las bajas que vamos a ir relatando. Desaparecieron las cuatro familias de tortugas marinas (aunque tres de las cuatro sobrevivieron en otros lugares). De las tres familias de crocodílidos, una, la Goniopholidae, también cayó. Lo mismo sucedió con dos familias de pterosaurios (reptiles voladores), la Ornithocheiridae y una familia todavía sin denominar, otras dos de ictiosaurios (reptiles marinos), la Platypterygiidae y otra familia todavía innominada, así como las tres familias de plesiosaurios (reptiles marinos también), la Elasmosauridae, la Polycotylidae y la Cimoliasauridae. Las ocho familias de lagartos perdieron dos: Mosasauridae (grandes formas marinas) y Polyglyphanodontidae (primitivas formas terrestres como el actual *Scincus* o pez de la arena, del norte de Africa, fusiforme y adaptado a enterrarse en la arena. N. del T.).

¿Qué ocurrió? ¿Se trató de una extinción gradual o de una catástrofe repentina? Mi área de investigación se centra en los grandes reptiles norteamericanos del Mesozoico. Ello explica por qué el análisis de estas cuestiones se ocupará principalmente de la desaparición de tales vertebrados. Entre las muchas hipótesis avanzadas para dar razón de su pérdida las hay que defienden una ruptura de la cadena alimentaria, tanto en el mar como en tierra firme, otras hablan de una alteración general del ambiente, en particular un descenso del nivel del mar a finales del Mesozoico, éstas un brusco aumento de la temperatura, aquéllas un descenso de la temperatura causada por polvo volcánico en la atmósfera, etcétera. Ninguno de estos fenómenos, sin embargo, parece constituir por sí solo una causa convincente de la extinción de los reptiles.

En 1979 se abrió una nueva vía a los paleontólogos interesados en el problema. Un grupo de investigadores de la Universidad de California en Berkeley (el geológo Walter Alvarez, su padre, el físico Luis W. Alvarez, y dos químico-físicos, Frank Asaro y Helen

EL NIVEL DE ARCILLA de la fotografía de la página opuesta tiene dos centímetros de potencia. Separa dos capas de calizas marinas que afloran cerca de Gubbio, en los Apeninos italianos. La caliza blanca que queda por debajo de la arcilla es del Mesozoico tardío y la caliza rosa agrisada superior es de principios del Cenozoico. El análisis de la arcilla mostró que era 30 veces más rica en el elemento pesado iridio, depositado en la superficie terrestre en material meteorítico, que las arcillas de estratos marinos adyacentes. Esto ha llevado al geólogo Walter Alvarez, a su padre, el físico Luis W. Alvarez, y a dos químico-físicos, Frank Asaro y Helen V. Michel, de la Universidad de California en Berkeley, a emitir la hipótesis de que el exceso de iridio procede de un objeto extraterrestre, quizás un meteorito del tamaño de un asteroide, que chocó contra la Tierra a finales del Mesozoico. El grupo de Alvarez sugiere, además, que la colisión fue la causa de las numerosas extinciones de organismos terrestres y marinos que tuvieron lugar en esa época. La moneda que aparece sobre el estrato de arcilla es una pieza italiana de 50 liras.

SEDIMENTOS EROSIONADOS del Cretácico tardío, en las inmediaciones de Fort Peck Reservoir, Montana. Tienen 100 metros de potencia. Se depositaron en una llanura costera cercana a la orilla de un gran mar interior que existió allí durante el Mesozoico. A la izquierda del hombre que aparece en primer plano hay un cráneo del dinosaurio *Triceratops*, aflorando parcialmente por meteorización. Los últimos restos esqueléticos articulados de dinosaurios se encuentran en este nivel, en la parte alta de la formación. La figura de blanco (*arriba*) se halla sobre una cata vertical que corta un estrato oscuro poco potente. Una segunda cata, a la izquierda, se extiende sobre el mismo estrato. En este nivel, 5,5 metros por encima del cráneo de *Triceratops*, el polen y las esporas fósiles dan prueba de que las plantas empezaron a escasear de golpe. Las extinciones de plantas y reptiles fueron casi simultáneas.



TERCER AFLORAMIENTO de sedimentos en el límite entre el Mesozoico tardío y el Cenozoico, cerca de Aix-en-Provence, Francia. El límite está cerca del nivel a media distancia, donde se aprecia al autor del artículo excavando una cata. Estos sedimentos no son de origen marino, sino continental.

V. Michel) anunciaron el descubrimiento de trazas anormalmente grandes del elemento pesado iridio en una formación marina próxima a Gubbio, en los Apeninos italianos. El iridio se concentraba en un nivel de arcilla, de uno a dos centímetros de potencia, que separaba unas calizas marinas de finales del Cretácico de otras calizas marinas, suprayacentes, de principios del Paleoceno. La caliza situada por debajo de la arcilla contenía organismos fósiles marinos típicos de las postrimerías del Cretácico. No había restos de organismos en la arcilla. En la caliza situada por encima de la arcilla tampoco aparecieron organismos cretácicos: habían sido sustituidos por otros, típicos del Paleoceno.

El iridio se inscribe en el grupo de elementos que los geólogos llaman siderófilos. Muy raros en rocas de la corteza terrestre, abundan bastante, en cambio, en los meteoritos. La constante lluvia de micrometeoritos sobre la superficie de la Tierra (de los cuales más del 70 por ciento caen en los océanos) aporta modestas concentraciones de iridio y otros elementos siderófilos en los sedimentos que se acumulan en las cuencas oceánicas.

En 1977, Walter Alvarez participaba en un trabajo de equipo, integrado por un grupo internacional de especialistas, en el examen de los estratos marinos cercanos a Gubbio y, entre otros, el nivel de arcilla. (Se encontraba allí también la paleontóloga Isabella Premoli Silva, de la Universidad de Milán.) Dado que la caída de material micrometeorítico se considera más o menos constante, Luis Alvarez sugirió que, midiendo la cantidad de iridio en la arcilla, podría calcularse cuánto tiempo había tardado en depositarse el nivel. Cuando Asaro y Michel lo midieron, descubrieron con sorpresa que el iridio contenido en el nivel de arcilla era 30 veces más abundante que en las arcillas de los estratos de caliza adyacentes.

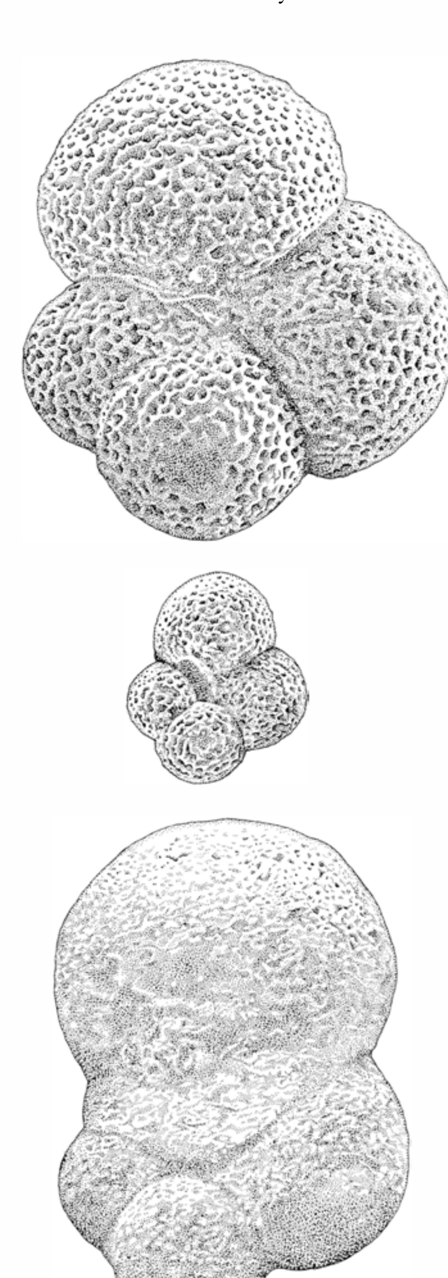
Si este exceso de iridio hubiera procedido, de la forma que fuere, de fuentes terrestres, la arcilla hubiera mostrado enrarecimientos comparables de otros elementos normalmente asociados a los minerales que integran la arcilla. El análisis del grupo de Berkeley reveló un patrón distinto de enrarecimiento, más cercano al de los porcentajes relativos de los elementos encontrados en los meteoritos. ¿Podría el exceso de iridio proceder de un depósito oceánico de elementos derivados de los micrometeoritos, repentinamente precipitados por un acontecimiento químico? Evidentemente no; ni por encima ni por debajo del nivel de arcilla había señal alguna de que la proporción normal de acumulación de siderófilos hubiera descendido, como correspondería a una tal precipitación. En conexión con este hecho, Charles J. Orth, del Los Alamos Scientific Laboratory, y sus colaboradores han descubierto un exceso de iridio en la última capa de materiales sedimentarios del Cretácico de origen continental. La precipitación a partir de un depósito oceánico no explica, desde luego, su aparición.

Cuando se publicó el informe, en 1980, el grupo de Berkeley estaba en condiciones de añadir, a los datos de Gubbio, el descubrimiento de estratos portadores de iridio en exceso en sedimentos marinos del Cretácico tardío de Dinamarca, España y Nueva Zelanda. (Más tarde, agregarían a esa lista anomalías de iridio en testigos submarinos profundos del Atlántico y el Pacífico.) El informe terminaba con la sospecha siguiente: coincidiendo con la extinción en masa de ciertos microorganismos marinos (y consecuentemente con la de muchos reptiles), medio billón de toneladas de material extraterrestre cayeron bruscamente sobre la superficie de nuestro planeta.

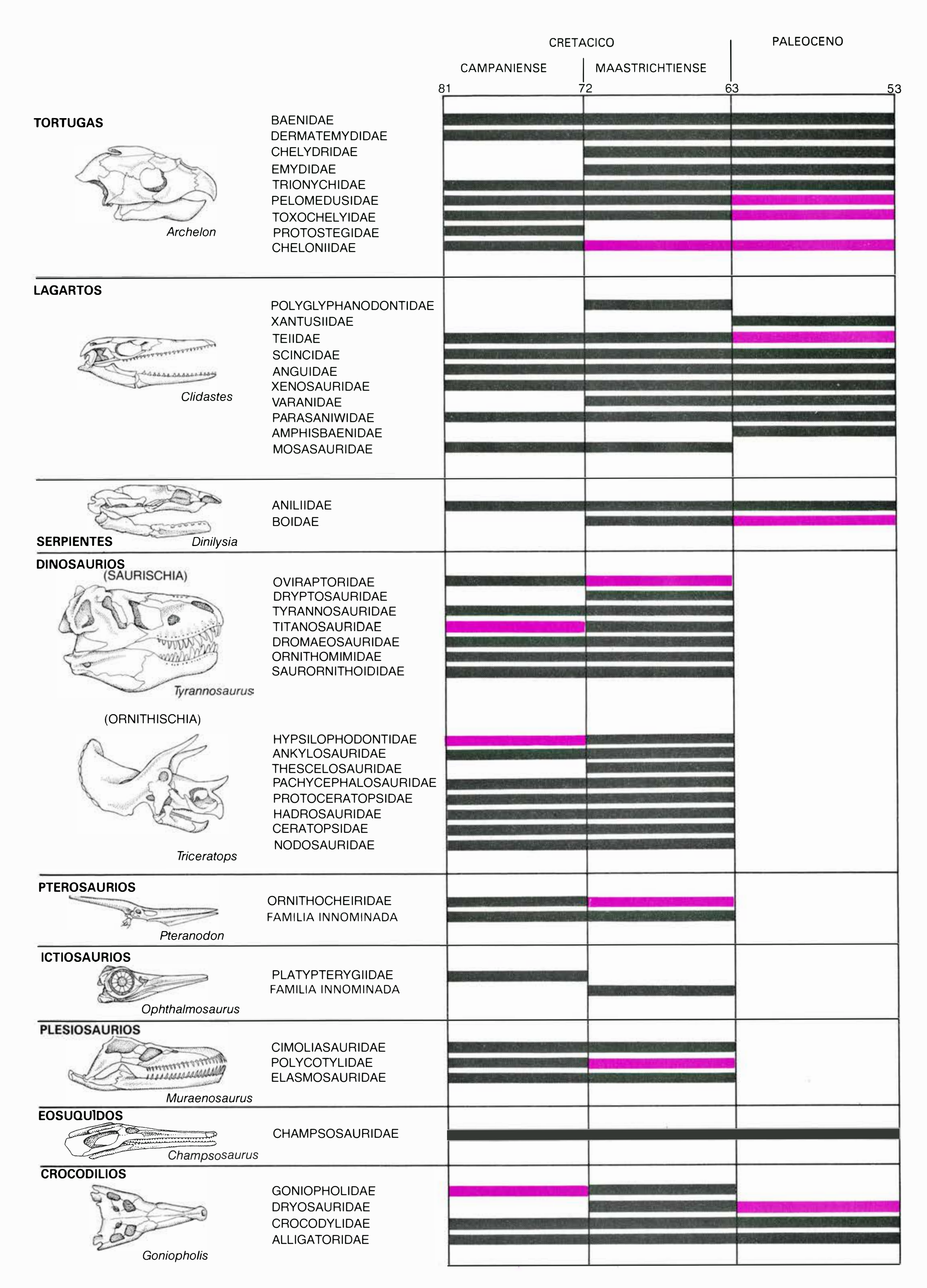
¿De dónde podía proceder semejante envío? La pregunta encaja en la elaboración consiguiente de la hipótesis del grupo de Berkeley. Si, por una parte, el material procedía del propio sistema solar, el mecanismo de su llegada a la Tierra podría hallarse en el gran volumen de datos que se dispone de esta región del espacio. Pero si venía de más lejos, donde escasean los datos relativos a múltiples ambientes, la propuesta de cualquier mecanismo carecería de base observacional.

Consideremos la posibilidad de que el material extraterrestre se produjera fuera del sistema solar por una explosión gigantesca: una supernova. En apoyo de tal hipótesis, Malvin A. Ruderman, de la Universidad de Columbia, y James W. Truran, Jr., de la de Illinois en Urbana-Champaign, sostienen que un brote gigante de rayos gamma de una explosión de ese tipo podría haber levantado material micrometeorítico de la superficie de la Luna, material que la Tierra había recogido. Pero matizan, a renglón seguido, que todavía están por ver los destellos de rayos gamma de las supernovas, amén de que la transferencia de iridio hasta la Tierra por semejante procedimiento no cubriría la cantidad demandada.

La hipótesis de la supernova espera recibir apoyo también del estudio en que está empeñado Paolo Maffei, del Observatorio Astrofísico de Catania, Italia: evaluación de las pruebas astronómicas de un hecho explosivo gigantesco, a unos 1000 años-luz del sistema solar, en las postrimerías del Mesozoico. Pero Wallace H. Tucker, del Centro de Astrofísica del Observatorio de la Facultad de Harvard y del Observa-



FORAMINIFEROS del Mesozoico final e inicios del Cenozoico en dibujos basados en micrografías de microscopio electrónico, representadas a la misma escala (200 aumentos). Miembros de los Globigerinacea, son, de arriba abajo, *Eoglobigerina*, un especimen de principios del Paleoceno de Dinamarca, un ejemplar de *Eoglobigerina* de los inicios del Paleoceno de un testigo profundo del Pacífico Norte y una *Rugoglobigerina* erosionada del Cretácico final de un testigo de un pozo petrolífero de Libia. El tamaño anormalmente pequeño del segundo ejemplar se considera reflejo de la recuperación gradual de los condicionamientos ambientales que causaron la desaparición mesozoica.



CRETACICO
PALEOCENO
CAMPANIENSE
MAASTRICHTIENSE
81
72
63
53
TORTUGAS
Archelon
BAENIDAE
DERMATEMYDIDAE
CHELYDRIDAE
EMYDIDAE
TRIONYCHIDAE
PELOMEDUSIDAE
TOXOCHELYIDAE
PROTOSTEGIDAE
CHELONIIDAE
LAGARTOS
Clidastes
POLYGLYPHANODONTIDAE
XANTUSIIDAE
TEIIDAE
SCINCIDAE
ANGUIDAE
XENOSAURIDAE
VARANIDAE
PARASANIWIDAE
AMPHISBAENIDAE
MOSASAURIDAE
SERPIENTES
Dinilysia
ANILIIDAE
BOIDAE
DINOSAURIOS
(SAURISCHIA)
Tyrannosaurus
OVIRAPTORIDAE
DRYPTOSAURIDAE
TYRANNOSAURIDAE
TITANOSAURIDAE
DROMAEOSAURIDAE
ORNITHOMIMIDAE
SAURORNITHOIDIDAE
(ORNITHISCHIA)
Triceratops
HYPSILOPHODONTIDAE
ANKYLOSAURIDAE
THESCELOSAURIDAE
PACHYCEPHALOSAURIDAE
PROTOCERATOPSIDAE
HADROSAURIDAE
CERATOPSIDAE
NODOSAURIDAE
PTEROSAURIOS
Pteranodon
ORNITHOCHEIRIDAE
FAMILIA INNOMINADA
ICTIOSAURIOS
Ophthalmosaurus
PLATYPTERYGIIDAE
FAMILIA INNOMINADA
PLESIOSAURIOS
Muraenosaurus
CIMOLIASAURIDAE
POLYCOTYLIDAE
ELASMOSAURIDAE
EOSUQUIDOS
Champsosaurus
CHAMPSOSAURIDAE
CROCODILIOS
Goniopholis
GONIOPHOLIDAE
DRYOSAURIDAE
CROCODYLIDAE
ALLIGATORIDAE

torio Astrofísico Smithsoniano, considera esta hipótesis poco verosímil. El material interestelar barrido por una supernova, afirma, aunque muy grande, no se acumularía en concentraciones suficientemente importantes como para alcanzar la cantidad de iridio contenida en la zona de enrarecimiento siderófilo que cubre los estratos del Mesozoico tardío.

La secuencia de acontecimientos que se desencadenan en la explosión de una supernova empieza por una implosión inicial. En el transcurso de este colapso, los núcleos de los elementos pesados del centro estelar capturan rápidamente neutrones. Entre las nuevas especies nucleares que se forman está el plutonio 244. Viene luego una explosión, que distribuirá ese isótopo radiactivo a través de un enorme volumen de espacio. Los químicos del grupo de Berkeley buscaron Pu-244 en la arcilla rica en iridio, razonando que, si el iridio había sido producido por una supernova, debería haber plutonio en cantidades detectables. No hallaron nada en el rastreo.

Argumentaron luego que los dos isótopos del iridio, el Ir-191 y el Ir-193, se producirían, en proporciones distintas, en diferentes explosiones de supernovas, dadas las variaciones particulares de los flujos de neutrones y de los tiempos de reacción. Cuando analizaron la arcilla rica en iridio, encontraron que los dos isótopos se presentaban en la proporción que corresponde típicamente al iridio del sistema solar. Lo mismo ocurría con la proporción de los dos isótopos del osmio, como descubrieron, por separado, J. Hertogen, de la Universidad de Lovaina, y Ramachandran Ganapathy, de la J. T. Baker Chemical Company de Phillipsburg, New Jersey. Por tanto, parece improbable que el material extraterrestre presente en el nivel de arcilla fuera el producto de la explosión de una supernova (distinta de la que pudo ser responsable de la formación del sistema solar).

Si el material procedía del sistema solar, ¿cómo alcanzó la superficie de la Tierra? Dos posibilidades surgen inmediatas: el choque contra un meteorito del tamaño de un asteroide o el choque contra un cometa. Respecto a la primera posibilidad, el grupo de Berkeley estima que la cantidad de material extraño de la capa rica en iridio esparcido por todo el mundo podría haber estado contenida en un asteroide de unos 10 kilómetros de diámetro. Pero la hipótesis del asteroide tiene un punto débil: la concentración de materiales de origen terrestre en la capa rica en iridio no es la que cabría esperar si hubieran sido arrancados de la corteza por el impacto de un cuerpo de tales dimensiones. Richard A. F. Grieve, del Ministerio canadiense de Energía, Minas y Recursos, ha sugerido un camino para salvar el obstáculo: el iridio podría haberse depositado, después del impacto, en forma de caída de material meteorítico, relativamente puro, reinyectado en la estratosfera por la fuerza de la colisión. Tanto el grupo de Berkeley como Jan Smit, de la Universidad de Amsterdam, geólogo que ha estudiado los estratos de caliza de fines del Mesozoico en España, ven otra vía para el problema. Si el asteroide cayó en el mar, que entra dentro de lo estadísticamente posible, sólo una pequeña cantidad de material de la corteza habría sido excavado por el impacto.

¿Qué decir del choque contra un cometa? Estos cuerpos poseen baja densidad y están formados en gran parte por hielo. Se calcula, pues, que un cometa que contuviera siderófilos en cantidad suficiente para explicar el enriquecimiento observado debería tener una masa doble que la del hipotético asteroide y, por ende, un tamaño mucho mayor. Lo que, de nuevo, pone sobre el tapete el problema de la excavación. Para orillar la dificultad, Frank Kyte, de la Universidad de California en Los Angeles, y sus colegas Zhiming Zhou y John Wasson proponen que, cuando el cometa en cuestión se acercaba a la Tierra, debió desintegrarse a causa de las fuerzas gravitatorias. La Tierra habría recibido una lluvia de restos cometarios que no tenían por qué excavar ningún cráter notable u otro tipo de boquetes afines.

Cualquier acontecimiento de los que barajan las distintas hipótesis podría causar condicionamientos extremos, de corta duración, en el seno de la biosfera. Así S. V. M. Clube y William M. Napier, del Real Observatorio de Edimburgo, estiman que la onda de choque generada por el impacto en tierra de un asteroide tan grande no destruiría tan sólo todos los bosques de la Tierra sino que también mataría todos los grandes animales que poblaran los continentes. Y si el impacto fuera en el océano, generaría olas mareales de ocho kilómetros de altura.

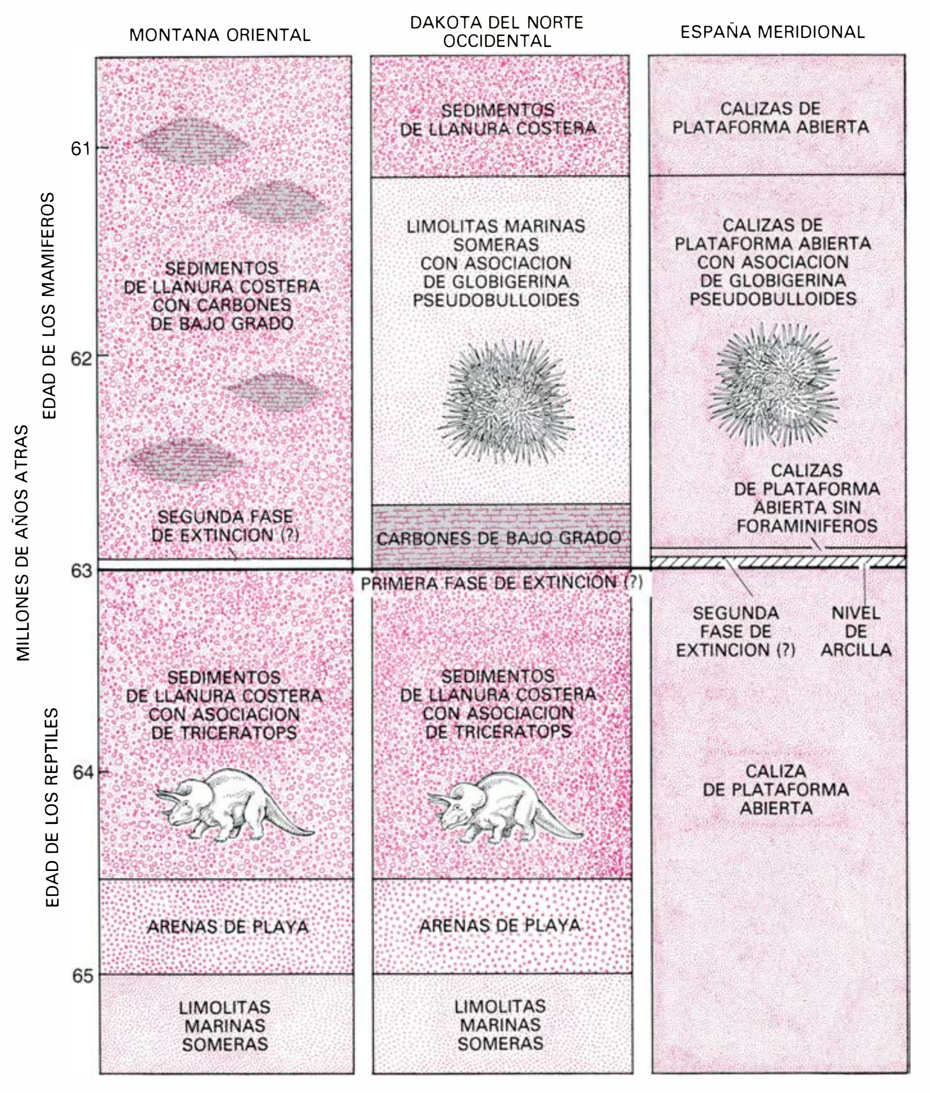
En opinión de Walter Alvarez y los demás miembros de su grupo, situaciones tan extremadas quizá no bastaran para provocar las extinciones de finales del Mesozoico. Y sugieren que el impacto habría inyectado, también, una enorme cantidad de partículas, del tamaño del polvo, en la estratosfera. Estas, enrarecerían la transparencia atmosférica y cuartearían los cimientos de la biosfera al disminuir la fotosíntesis.

¿Qué luz arroja el registro fósil sobre esta nueva hipótesis? No poca. Aunque debe recordarse que el registro fósil contiene una cantidad limitada de información que, además, no es fácil interpretar. Así, aunque los paleontólogos han venido desenterrando restos de animales mesozoicos desde hace más de un siglo, el número total de fragmentos de esqueletos de dinosaurios conocidos apenas ronda los 5000. Ello obedece, en parte nada despreciable, a puras razones crematísticas: resulta labor muy cara la de recuperar huesos de dinosaurio. Para que el lector se haga una idea, piense que el estudio del polen de plantas mesozoicas ofrece abundantes datos sobre el ambiente. Pero extraer unos 20.000 granos de polen y montarlos para su observación microscópica viene a costar unos 500 dólares (incluido un período de trabajo de una semana). Recoger y preparar el mismo número de huesos de dinosaurio valdría unos 400 millones de dólares (incluido un período de trabajo de un millón de semanas).

La comparación podría extenderse a un factor importante en el estudio de las extinciones biológicas: el tamaño de la muestra. Uno de los yacimientos de grupos de dinosaurios mesozoicos más recientes, los restos de animales que vagaban por las llanuras interiores de Estados Unidos y Canadá hace unos 63 millones de años, se caracteriza por la presencia de herbívoros de gran cornamenta, como el género *Triceratops*, y carnívoros gigantes, como los del género *Tyrannosaurus*. Los estratos, 12 millones de años más antiguos, del Dinosaur Provincial Park de Alberta (Canadá), han proporcionado una diversidad de dinosaurios mucho más rica que los materiales más recientes.

¿Significa ello, como algunos han sugerido, que los dinosaurios estaban reduciendo el número de especies a medi-

EXTINCION DE LOS REPTILES a fines del período Cretácico; se representa en la tabla de la página opuesta. De las 48 familias citadas, 24 desaparecieron aparentemente en las postrimerías del último piso del Cretácico, el Maastrichtiense, hace unos 63 millones de años. Una prolongación coloreada de la línea negra en el Cretácico indica un registro de la familia sólo en lugares no situados al norte de Río Grande. Una línea coloreada en la columna del Paleoceno indica la supervivencia de la familia fuera de Norteamérica. Con excepción de la serpiente, los demás cráneos mostrados son representativos de especies extinguidas. Los dinosaurios desaparecieron por completo y no hubo orden de reptil que no se viera afectado.



SECUENCIAS de dos localidades terrestres norteamericanas y una marina española durante el final del Mesozoico y el principio del Cenozoico. La presencia del mismo foraminífero reemplazante, *Globigerina pseudobulloides*, en Dakota del Norte y España destaca la coincidencia de extinciones terrestres y marinas.

da que el Mesozoico tocaba a su fin? ¿Habían iniciado su extinción millones de años antes de la presunta catástrofe? En absoluto. Los fósiles de Alberta muestran mayor variedad porque allí se han recogido muchos más ejemplares: más de 300 en comparación con el número escaso, siempre inferior a 75, cuando no seis o siete especímenes, de las localidades más recientes.

La realidad es que, por lo que concierne a Europa, ésta mantuvo casi incólume su diversidad a medida que concluía la era Mesozoica. Aumentó en Mongolia. En otras áreas del mundo las muestras son demasiado pequeñas para traslucir tendencias. No hay pruebas, en resumen, de que se asistiera a una disminución a largo plazo de la diversidad de los dinosaurios antes de su extinción.

En ningún otro punto del planeta se ha llevado a cabo un rastreo del registro fósil de organismos terrestres, de los últimos 1,75 millones de años del Cretácico, tan exhaustivo como el acometido en los afloramientos de sedimentos marrón agrisados, de unos 100 metros de espesor, situados alrededor del borde sur del Fort Peck Reservoir, al nordeste de Montana. En la mitad inferior, más antigua, de la formación predominan las arenas de río; en la superior, más moderna, las arcillas y limos aluviales. El cambio de modelos deposicionales refleja presumiblemente cambios ambientales que podrían haber causado una alteración en la distribución de los animales que poblaban lo que fue una llanura costera, dinosaurios incluidos. En los niveles inferiores predominan el género depredador *Tyrannosaurus*, de gran tamaño, y el género herbívoro con pico de pato *Amatosaurus*. Pero en los niveles superiores los que más abundan son el género herbívoro *Triceratops* y pequeños dinosaurios ramoneadores del género *Thescelosaurus*. Cambios similares en la comunidad de plantas están documentados por estudios de las esporas y pólenes fósiles realizados por Robert Tschudy, del Servicio Geológico de los Estados Unidos.

Un extraordinario modelo de cambio aparece en el techo de la formación de 100 metros. Aquí, a lo largo de un horizonte, las abundancias relativas de la secuencia inferior se invierten y los restos de grandes dinosaurios sobrepasan localmente a los de los pequeños. Los grandes animales suelen caracterizarse por tasas de natalidad y mortandad menores que las correspondientes a los animales pequeños. De ahí que una proporción de este tipo, más dinosaurios grandes que pequeños conservados en estado fósil, sugiera algún tipo de muerte en masa. (Mi interpretación es provisional, pero futuros trabajos en Fort Peck Reservoir clarificarán seguramente la situación.)

Por encima de este horizonte, inmediato al techo de la formación, los sedimentos adquieren un aspecto más laminado. Los únicos huesos de dinosaurios son unos cuantos fragmentos que parecen haber sido erosionados a partir de estratos más antiguos y depositados secundariamente en lechos de corrientes de agua. En principio no hay cambios equivalentes en la comunidad vegetal. Luego, en un nivel sito unos cinco metros por encima del horizonte que contiene los últimos materiales esqueléticos de dinosaurios no redepositados, apenas si se conserva polen y esporas fósiles. David M. Jarzen, del Museo Nacional de Ciencias Naturales de Ottawa, ha estudiado las pruebas vegetales. Ambos creemos que el intervalo de cinco metros representa un período deposicional de varias decenas de miles de años de duración.

Por encima del nivel de magro registro fósil vegetal, capas de carbón de bajo grado alternan con limolitas laminadas. Esta separación entre la nutrida presencia de dinosaurios y la muy rala de carbón la observó, por primera vez, hace algunos años William A. Clemens, de la Universidad de California en Berkeley. Tschudy, que ha estudiado también el polen y las esporas de los estratos con carbón, indica que estas pruebas de vida vegetal son tres veces menos abundantes que las de estratos inferiores, con huesos de dinosaurios. La edad de los mamíferos había empezado.

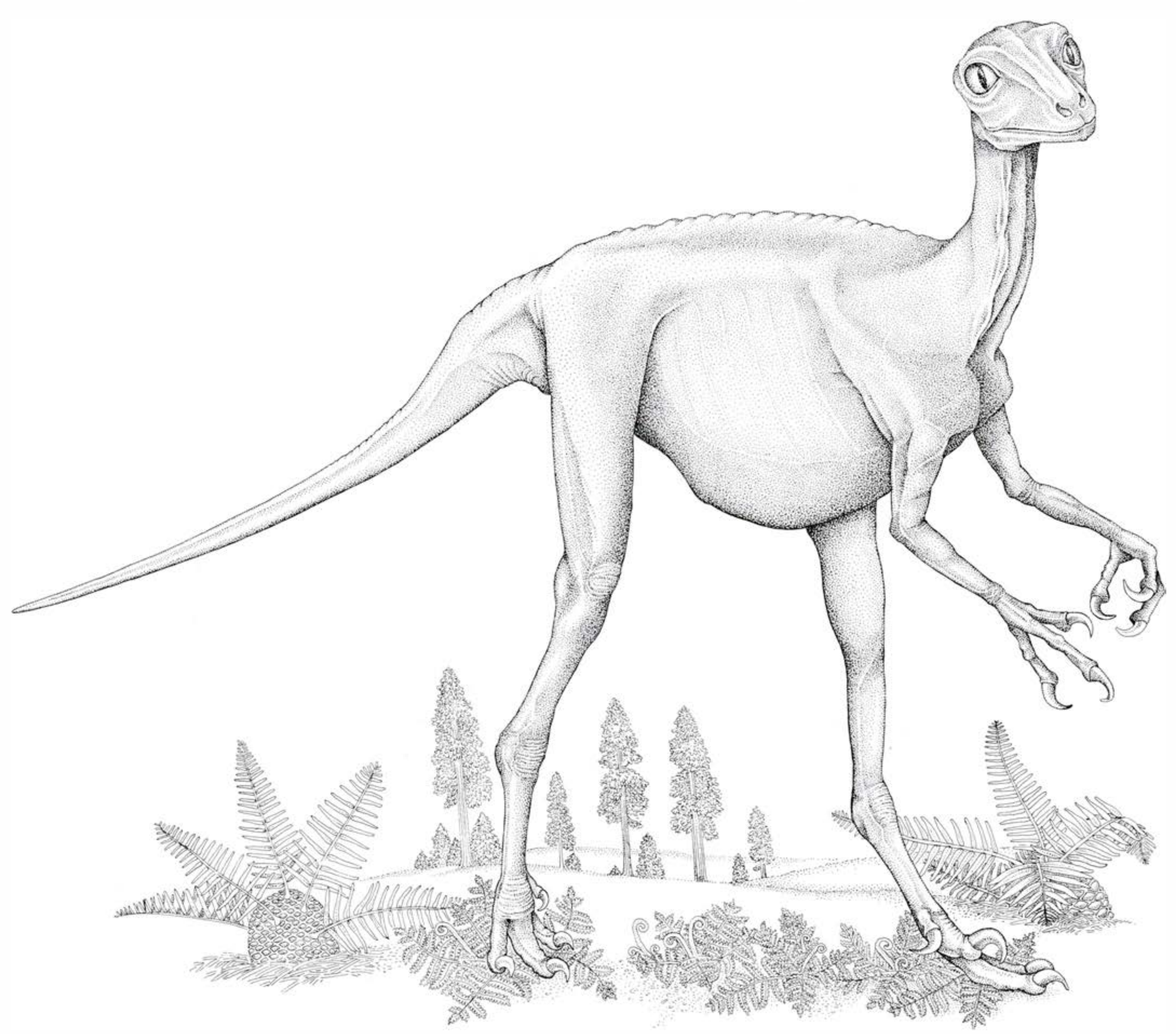
Las calizas que Smit y sus colegas han estudiado en España registran una sucesión de acontecimientos muy peculiar. Se depositaron en el fondo de un mar tropical abierto, que se extendía

por el sur de España a finales del Mesozoico. Las formaciones están compuestas, de un modo casi exclusivo, de caparazones y plaquitas de carbonato cálcico de pequeños foraminíferos: protozoos de libre flotación que pertenecían al zooplancton marino. Durante más de 10 millones de años, la productividad planctónica del lugar debió ser alta y no hubo cambios significativos en el carácter de los restos orgánicos depositados en el fondo marino.

Luego, en un nivel de roca de no más de cinco milímetros de potencia (que viene a representar menos de 200 años de deposición), desaparece casi el 90 por ciento de las especies de foraminíferos encontrados por debajo del mismo. Los protozoos supervivientes alcanzaron sólo una décima parte del tamaño de sus predecesores. A medida que cesó la lluvia de caparazones y plaquitas también lo hizo la acción bioturbadora de los invertebrados marinos que poblaban el fondo. Un manto de arcillas laminadas rojas y grises se acumuló en el suelo marino, alcanzando una potencia de unos 10 centímetros. Las condiciones permanecieron aparentemente estables durante, quizás, unos 20.000 años; por aquel entonces, todas las especies de foraminíferos supervivientes, salvo una, habían ido declinando hasta acabar extinguiéndose.

Cerrado ese intervalo, la vida empezó a proliferar una·vez más. Reanudóse la deposición de sedimentos; el fondo marino volvió a sentir la presencia de invertebrados bentónicos. Apareció una nueva asociación de foraminíferos, a la que pronto siguió otra, caracterizada por la presencia, entre varias, de la especie *Globigerina pseudobulloides*. Los protozoos resurgidos poblaron el antiguo mar español a lo largo de los siguientes dos millones de años.

A medio mundo de distancia, en lo que hoy es Dakota del Norte, un gran mar interior se expandía hacia el oeste por esas mismas fechas, inundando una área deltaica donde los restos de *Triceratops* se fosilizaron y quedaron cubiertos por estratos con carbón. Las limolitas marinas que se depositaron sobre el carbón contienen caparazones de especies de foraminíferos pertenecientes a la misma asociación de *G. pseudobulloides* que aparece en el mar español tras la extinción de los grandes foraminíferos. Dada la gran inseguridad para evaluar, a partir de la potencia de depósitos sedimentarios, el tiempo transcurrido, parece posible que la historia relatada por los sedimentos aquí, en Fort Peck Reservoir, y en el sur de España sea la misma. En caso afirmativo,



EL DINOSAURIO DE GRAN CEREBRO del Cretácico tardío, *Stenonychosaurus inequalus*, medía unos tres metros del extremo del hocico al final de la cola. El dibujo se basa en una restauración del Museo Nacional de Ciencias Naturales de Ottawa, Canadá. Su coeficiente peso cerebral/peso corporal iguala el de los primeros mamíferos. Si tales depredadores hubieran sobrevivido, puede que hubieran obstaculizado el desarrollo de los mamíferos.

la extinción de los dinosaurios sobre la tierra y la de los foraminíferos en el mar habrían coincidido.

Los foraminíferos no fueron los únicos organismos marinos que desaparecieron a finales del Mesozoico. Mencioné antes el caso de bastantes reptiles marinos. También cayeron varios moluscos: los cefalópodos de concha enrollada conocidos como ammonites, los cefalópodos parecidos al calamar denominados belemnites y unos peculiares bivalvos coraliformes, los rudistas. La mayor parte de las principales familias de animales marinos sobrevivieron, pero perdieron muchos géneros y especies.

El registro fósil en este límite crucial se entiende mejor para la microfauna, como los foraminíferos, que para la macrofauna marina. La razón estriba en que la macrofauna es numerosa y diversa y, finito, el número de paleontólogos. Incluso en formaciones que podríamos considerar bien estudiadas, así las cretas de Dinamarca, la tasa de supervivencia entre grupos animales tan importantes como las esponjas, los braquiópodos, los gasterópodos marinos y los crustáceos está todavía por calcular. Tampoco se conoce bien el registro de la vida animal en las regiones tropicales del globo que se desarrolló por entonces. Ante tamaña parquedad de datos, a nadie extrañará que la polémica entre gradualistas y catastrofistas mantenga plena su viveza.

Una clara distribución en tablas resultará útil para calibrar la magnitud de las extinciones. Compárese el número de géneros animales en el registro fósil unos 10 millones de años antes del final del Cretácico con el número de géneros de un período equivalente después de la crisis. No vale la pena ponderar la provisionalidad de esos valores numéricos. A pesar de lo cual, las cifras reflejan un declive del 50 por ciento en la diversidad mundial de géneros. Cuando uno repite este juego de números, contando las especies registradas para determinados géneros de animales y vegetales, antes y después de la crisis, el resultado es similar. En una muestra que incluyera mamíferos en representación de los animales terrestres, algas quitinosas marinas por los vegetales marinos y erizos de mar, estrellas de mar y ostras por los animales marinos, el declive de especies durante el intervalo de extinción sería de unas tres especies por género a 1,5. Por tanto, parece razonable pensar que la crisis biológica asociada con la extinción de los dinosaurios también causó la desaparición del 75 por ciento de las especies animales y vegetales preexistentes. Y, para decir toda la verdad, presumo que se trata de una estimación un tanto moderada.

El inventario de extinciones pone al descubierto algunas anomalías. Por ejemplo, no sobrevivió ningún animal terrestre de más de 25 kilogramos; muchos de los que desaparecieron pesaban, además, bastante menos. También las plantas terrestres de las regiones septentrionales de la zona templada sufrieron más pérdidas que las meridionales. En cambio, animales y vegetales de las comunidades de agua dulce resultaron escasamente afectados. Lo mismo debió ocurrir a los moluscos marinos de aguas profundas, en opinión de Arthur H. Clarke, de Ecosearch Inc. de Mattapoisett, Massachusetts. La vida marina de aguas someras, sin embargo, especialmente la fauna de arrecifes tropicales, se alteró profundamente.

Ni siquiera los animales que compartían el mismo medio resultaron afectados por igual. Como ha señalado Eric Buffetaut, de la Universidad de París, los crocodilios que ocupaban aguas marinas someras sobrevivieron a las extinciones, pero los mosasaurios, que ocupaban el mismo hábitat, no. Cualesquiera que fueran los agentes de stress biológico, incluidas las perturbaciones en la cadena alimentaria, la capacidad de la biosfera para resistirlas fue, evidentemente, diversa.

¿Cuál es el significado de la naturaleza, aparentemente dual, de las extinciones de animales y vegetales a finales del Mesozoico? ¿Fueron las extinciones hechos realmente separados, esto es, desaparecieron primero los animales terrestres y luego las plantas? En caso afirmativo, ¿fue la segunda extinción resultado de condiciones tan severas como las que habían causado la primera, o se trató de un simple fenómeno biológico de carácter cuasi secuencial? Cualquiera que sea la respuesta que se dé a estas preguntas, el género humano puede haber sido el beneficiario a largo plazo del evidente cataclismo. Al acercarse el final del Mesozoico, ciertos microdinosaurios carnívoros habían alcanzado un coeficiente peso cerebral/peso corporal característico de los primeros mamíferos. Si estos reptiles, presumiblemente más inteligentes, hubieran sobrevivido, sus descendientes podrían haber continuado bloqueando el desarrollo de los mamíferos, eliminando así de antemano nuestra propia aparición y la situación privilegiada de organismos dotados del mayor desarrollo cerebral que conocen los biotas del planeta.

Ciencia y sociedad

El primer jardín botánico de España

Desde que Millás Vallicrosa dedicó una serie de monografías al estudio de la escuela agronómica hispanoárabe, conocemos bastante bien el interés que los estudios de esta índole adquirieron en la España Musulmana, por lo menos desde el siglo XI. La importancia de esta tradición repercutió directamente en la España Cristiana ya que, en el siglo XIII, se llevaron a cabo traducciones castellanas de los tratados de agricultura escritos por los toledanos Ibn Wáfid e Ibn Bassal (ambos del s. XI) y cuando en pleno Renacimiento Gabriel Alonso de Herrera escribe su *Agricultura General*, cita repetidamente la obra de Ibn Wáfid. En un estudio de conjunto sobre la escuela agronómica andalusí, publicado por Lucie Bolens en 1974, se señalaba que el nivel de conocimientos alcanzado por estos tratados hispanoárabes no sería superado hasta el siglo XIX y esto nos explica el que el ministro Campomanes justificara, en 1793, la reedición del texto árabe y traducción castellana de la *Agricultura* de Ibn al-Awwam (debida a J. A. Banqueri), con las siguientes palabras que aparecen en el prólogo de la obra: "El crédito del presente tratado hizo desearle en nuestro idioma en el reinado del señor Fernando VI de augusta memoria; habiéndose impreso en el año 1751, traducidos en castellano, los capítulos XVII y XIX, que son el primero y el tercero de la segunda parte, con notas y un prólogo en que se recomienda la utilidad de esta obra y la ventaja que sacaría la nación de que se tradujese e imprimiese íntegra, yo extendí en aquel tiempo este prólogo, notas y versión castellana y desde entonces permanezco en el dictamen de que el tratado de Ebn el Awam no sólo es provechoso sino absolutamente necesario para mejorar la labranza y la crianza en España". Este párrafo de Campomanes, que ha sido puesto de relieve repetidamente por Vernet, muestra cómo, a fines del siglo XVIII, se publica una versión castellana de un tratado de agronomía árabe no por interés puro erudito, sino porque se le considera útil para el desarrollo de la agricultura en la España del tiempo.

Uno de los factores comunes más claros en esta escuela agronómica andalusí es el recurso continuo a la experiencia. Desde el siglo XI abundan las referencias al interés que se sentía por el estudio de las plantas procedentes de otras latitudes y a los ensayos que se realizaban para aclimatarlas a nuestro país. A fines del siglo XI o principios del XII un interesantísimo tratado de botánica, anónimo y conservado en un manuscrito de la Real Academia de la Historia, en el que se encuentra un primer intento de clasificación taxonómica de los vegetales, nos da información sobre plantas exóticas procedentes de la India, China, isla de Socotora, Yemen, Arabia y Jurasán: el autor recoge sus datos de drogueros ambulantes dedicados al comercio de plantas medicinales de un país al otro. Sabemos que este mismo botánico anónimo realizó ensayos, tanto con plantas indígenas como con plantas exóticas, en el Jardín del Sultán en Sevilla y en su propio huerto de esta ciudad. Este autor nos refiere asimismo que su maestro Ibn al-Luengo le enseñó semillas de mirobalano índico que había tomado de unas cuantas que poseía su maestro, el antes citado médico y agrónomo toledano Ibn Wáfid (m. 1075): estas semillas le habían sido traídas a al-Mamún, rey de la taifa de Toledo (1043-1075), desde la India Septentrional. Millás señaló también la importancia de los cultivos experimentales en el siglo XI y adujo referencias al cultivo del granado salvaje y al de una higuera cuyo fruto era mitad verde y mitad blanco. En esta época sabemos que existieron, por lo menos, dos jardines botánicos: uno en Toledo, a orillas del Tajo, y otro en Sevilla, al que ya he aludido.

Ahora bien, no resulta imposible intuir la existencia de antecedentes de estas experiencias e instituciones en los siglos anteriores al XI: Millás creía que en la Córdoba Califal, o sea antes del 1031, existieron ya jardines botánicos o campos de experimentación, destinados a la aclimatación de plantas procedentes del Medio Oriente. Nada tendría esto de particular, ya que hay motivos para pensar que la escuela agronómica andalusí del siglo XI deriva, posiblemente, del magisterio indirecto del gran médico y cirujano del siglo anterior Abulcasis (Abu-l-Qásim al-Zahrawi), quien sería asimismo el autor de un tratado de agricultura que conservamos en forma manuscrita. Pero incluso en el siglo IX podemos rastrear algunos datos sobre la existencia de una tradición geopónica de carácter experimental: hace ya bastantes años que García Gómez dio a conocer un curiosísimo texto árabe en el que se nos refiere el procedimiento utilizado por el poeta, astrólogo y diplomático Yahya al-Gazal para sacar subrepticiamente de Bizancio, metidas dentro de las páginas de un libro, con ocasión de su embajada a esta ciudad en el 840, semillas de una nueva variedad de higuera que fueron sembradas en Córdoba, logrando que dieran fruto. Los higos fueron presentados al emir, quien fue informado de la estratagema utilizada por al-Gazal y se mostró interesado por la experiencia. Vernet cree, por otra parte, que al-Gazal pudo también traerse de Bizancio gusanos de seda: la industria de la seda se había desarrollado en el Imperio Bizantino desde el siglo VI y las primeras alusiones a ella aparecen en España poco después del 840, mucho antes que en el resto de Europa.

Los que se han preocupado, hasta ahora, de la historia de la agronomía hispanoárabe y han recogido las referencias, a las que acabo de aludir, a jardines botánicos y cultivos experimentales no conocían, según parece, un interesantísimo pasaje del historiador Ibn Saíd (m. 1285) que contiene dos partes claramente diferenciadas: una primera en la que se nos dice textualmente que el primer omeya español, Abd al-Rahmán I (755-788) realizó un programa sistemático de aclimatación de plantas exóticas en los jardines que rodeaban su residencia de la Ruzafa, en Córdoba, jardines que se convierten, de este modo, en el primer jardín botánico conocido en la España Musulmana. La segunda parte, en cambio, está tomada del gran historiador cordobés Ibn Hayyán (987-1076), y nos refiere un ejemplo concreto de esta tarea: el éxito obtenido por el cortesano Safr b. Ubayd al-Kilaí, quien logró aclimatar, en una alquería malagueña, una nueva especie de granada que fue denominada, en lo sucesivo, granada *safarí*, en recuerdo de Safr, el hombre que la introdujo en España. El pasaje tiene, a mi entender, tanto interés que no puedo resistir la tentación de reproducir aquí una versión del mismo:

"De entre los parques de al-Andalus,

mencionaremos en primer lugar al que perteneció a los califas marwaníes, a saber el del alcázar de la Ruzafa. Refirió mi padre, Dios lo tenga en su gloria, lo siguiente: entre las construcciones que llevó a cabo Abd al-Rahmán I b. Muawiya, a principios de su reinado, se encuentra el parque de la Ruzafa, situado al noroeste de Córdoba, que quiso convertir en lugar de esparcimiento para pasar en él buena parte de su tiempo. Allí tuvo un hermoso palacio y situó amplios jardines a los que hizo traer plantas exóticas y magníficos árboles procedentes de las regiones más diversas. En ellos plantó los huesos de frutas seleccionadas y las semillas extrañas que le habían traído Yazid y Safr, sus embajadores en Siria, hasta que crecieron, en un breve espacio de tiempo, gracias a los serios esfuerzos y cuidados adecuados, árboles tocados con el turbante de sus hojas, que dieron curiosos frutos y se diseminaron por toda la tierra de al-Andalus. El monarca reconoció que estos frutos eran los mejores de su especie. Siguió diciendo mi padre: su abuelo Hisham fue quien acuñó la palabra Ruzafa, aplicándola a una zona de Siria que era su favorita. Abd al-Rahmán I le imitó al elegir el asentamiento de esta Ruzafa suya: se prendó de ella, la visitó con frecuencia y residió allí la mayor parte del tiempo. La fama de este lugar corrió en los días de Abd al-Rahmán I y sus sucesores conservaron la inclinación que él tenía por la Ruzafa. Mi padre dijo asimismo: todos sintieron predilección por ella y siguieron construyendo edificios allí, mientras los poetas la describían, rivalizando en esta tarea en poemas, los mejores de los cuales han sido transmitidos hasta nosotros".

Sigue el relato de Ibn Saíd: "La granada *safarí* que se difundió por todos los confines de al-Andalus, de tal modo que los habitantes de este país la preferían a todas las variedades restantes, tenía su origen en esta Ruzafa". Ibn Hayyán expone cuanto a ella se refiere: Esta variedad se distingue por su calidad, siendo la mejor de todas las especies de granada por su sabor dulce, su tacto suave, su carácter especialmente jugoso y la belleza de su forma. La trajo a al-Andalus el embajador enviado a Siria por Abd al-Rahmán I con el encargo de entrar en contacto con su hermana, el cual había traído los mejores ejemplares de granadas de la Ruzafa siria, cuya fundación se atribuye al califa Hisham. Abd al-Rahmán la mostró, entonces, a los hombres de su privanza, poniendo de relieve su belleza. Entre los presentes se encontraba Safr b. Ubayd al-Kilaí que era oriundo de Jordania. Se dice que era uno de los descendientes de aquellos habitantes de Medina que tomaron el partido del Profeta, Dios le bendiga y le salve, y que llevaban sus estandartes en sus campañas guerreras. Refiere también Ibn Hayyán que estos mismos descendientes eran los que llevaban las banderas delante de los califas omeyas. El monarca le entregó a Safr una parte de estas granadas con lo que él quedó maravillado ante su hermosura y quiso hacer una experiencia con ellas. Las llevó entonces a una alquería situada en la región de Málaga. Allí manipuló sus semillas y se las arregló para plantarlas, alimentarlas con agua y abono y transplantarlas hasta que surgió un árbol que dio fruto y éste, a su vez, maduró. Safr, entonces, arrancó los frutos de cuajo, quedando asombrado ante su belleza, y se dirigió con ellos inmediatamente a presencia de Abd al-Rahmán I, quien pudo comprobar que eran semejantes en todo a las granadas de la Ruzafa siria. El emir le preguntó cómo lo había conseguido y Safr le informó del procedimiento que había utilizado para obtenerlas. El monarca, entonces, admiró su descubrimiento, le dio las gracias por la tarea que había llevado a cabo y recompensó con generosidad su regalo. Acto seguido se plantó aquella



La botánica aplicada (farmacología) y la agronomía tuvieron un desarrollo brillante en la España Musulmana, sobre todo a partir del siglo x. En la ilustración, Dioscórides entrega la mandrágora a uno de sus discípulos: Dioscórides, para la botánica, y las ideas médicas de Hipócrates y Galeno, para la agronomía, son los dos fundamentos clásicos en los que se asienta el desarrollo de estos estudios en al-Andalus

granada en los jardines de la Ruzafa y en otros de su propiedad. Aquella especie de granada se difundió, el pueblo diseminó las plantaciones de esta índole y atribuyó su origen a Safr ya que, desde entonces y hasta ahora, es conocida como granada *safarí*."

El largo pasaje que acabo de traducir, tomándome con él ciertas libertades que pueden facilitar su comprensión, resulta suficientemente explícito y no requiere comentarios especiales. Sólo queda señalar que, en él, Abd al-Rahmán I, nostálgico de su patria siria, construye su palacio de la Ruzafa en Córdoba en recuerdo de la mansión y los jardines de su abuelo Hisham en Oriente. Si en la Ruzafa cordobesa se realizan ensayos de aclimatación de plantas orientales la causa es, una vez más, la nostalgia del monarca que pretende rodearse de las mismas plantas y frutas de su país de origen. No en vano será Abd al-Rahmán quien cante, de este manera, a una palmera que se alzaba en los jardines de su palacio: "¡Oh, palma! Tú eres, como yo, extranjera en Occidente, alejada de tu patria." (Julio Samsó.)

Carábidos del Pirineo aragonés

En la península ibérica, el abigarrado paisaje del Alto Aragón, con una notable variedad de climas locales y topoclimas del mayor interés, presenta condiciones inmejorables para estudiar las preferencias de residencia ecológica por parte de un notable número de especies de insectos que habitan también otras áreas geográficas de Eurasia o que son propias de la región (especies endémicas).

Entre los insectos coleópteros, comúnmente conocidos como escarabajos, los Carábidos, con unas 40.000 especies descritas, forman uno de los grupos más numerosos de insectos. Debido a sus exigencias concretas de condiciones ambientales y a su capacidad de reaccionar rápidamente ante una alteración del medio, así como por su gran independencia respecto a la vegetación –ya que en su mayor parte son carnívoros o en todo caso de alimentación variada–, los Carábidos son importantes bioindicadores de hábitats y de posibles interferencias humanas en los mismos. Además, su distribución espacial y temporal depende de factores abióticos que se relacionan preferentemente con la geomorfología, clima, microclima y tipo de cobertura vegetal.

Estos factores mencionados pueden reflejarse en una ecuación en la que intervienen ciertas variables climáticas

HABITAT	INDICE DE PRIMAVERA	SUBGRUPOS DE ESPECIES									
		1	2	3	4	5	6	7	8	9	10
Bajo piedras, en áreas descubiertas y praderas	Alto	++	+	++	+	++		++	+		
	Medio	++	+	++	++						
	Bajo	+	+		+		++				
Entre las piedras mojadas de pequeño tamaño que se encuentran en los bordes de ríos y arroyos (graveras)	Alto								+		
	Medio							+	++	++	
	Bajo							+	++	++	++

+ Poco frecuente; ++ Abundante.

Distribución espacial de los Carábidos considerando los distintos tipos de hábitat y la variación de los factores abióticos (reflejados en el IP)

(precipitación, temperatura), a la que hemos dado el nombre de "índice de primavera" –este índice se obtiene al restar la media, dividida por diez, de las temperaturas medias en grados Farenheit de los meses de primavera (marzo a mayo) de la media de los índices de aridez mensuales (precipitación en milímetros/temperatura media en grados Farenheit)–, que no sólo tiene una alta correlación con la distribución espacial y temporal de los insectos, sino que también se relaciona bien con los distintos tipos de paisaje. Así, al dominio del quejigal le corresponden índices bajos (entre −3,4 y −2,8); en estas áreas, el árbol potencialmente dominante sería el quejigo, roble de pequeñas hojas que caen en marzo y adaptado a la sequía. Junto a él son muy abundantes dos especies indicadoras de su degradación: el pino y el boj.

Al piso montano le corresponden índices medios (entre −2,8 y −1,5), entendiéndose como piso montano el área cubierta por bosques característicos de lugares húmedos: bosques de caducifolios (hayas, avellanos, etcétera), bosques de hayas con abetos, abetales y pinares húmedos con musgo y helechos. Este piso limita en su menor altitud con el dominio del quejigal y en su mayor altitud con los dominios subalpinos y de pastos de gran altitud.

A los dominios subalpinos y de los pastos de gran altitud les corresponden índices altos (Candanchú, por ejemplo, +0,3); entendemos como dominio subalpino el área donde se instala el pino negro acompañado de sotobosque de azalea de montaña y arándanos. Junto con los pastos de altitud ocupa las zonas más altas del Pirineo.

Las especies animales y vegetales no se distribuyen uniformemente sobre la superficie terrestre. Cada una de ellas ocupa un área de extensión variable en la que coexiste con otras especies, con las que comparte exigencias concretas de factores ambientales que se reflejan en el índice de primavera (IP). Por tanto, la composición local de la fauna depende de las preferencias de los distintos grupos de especies por uno u otro IP (es decir, por unas determinadas condiciones ambientales). Las especies eurosiberianas, que se distribuyen por Siberia, Europa septentrional y media y que llegan a veces a Africa del Norte, prefieren IP medios, tolerando bien los altos. Las especies mediterráneas, cuya distribución está delimitada en su borde norte por la presencia de cadenas montañosas próximas a la costa y en su parte africana y oriental llegan hasta las extensiones áridas del Sahara y de la región aralo-caspiana, prefieren los IP bajos. Las especies de Europa media y mediterránea, que además de ocupar el área mediterránea anteriormente citada se extienden más o menos ampliamente hacia el norte en la región eurosiberiana, prefieren índices de primavera medios, no tolerando los altos.

Los Carábidos pasan, a lo largo de su vida, por diversos estadios de desarrollo: huevo, larva, ninfa e imago (adulto sexualmente maduro y en general de corta vida). En IP altos, los Carábidos pasan al estadio de imagos durante los meses de julio y agosto; en IP medios durante los meses de junio, julio y agosto, y, en IP bajos, durante los meses de marzo y junio (disyunción temporal interesante para el estudio de los ciclos biológicos).

Al ser los Carábidos insectos que revelan gran tendencia a refugiarse bajo las piedras, el tipo de roca que aparece en un área determinada tiene un influjo significativo sobre su distribución: así, en granitos se reduce mucho el número de especies, mientras que en pizarras es rarísimo encontrar Carábidos, ya que los granitos al erosionarse originan bloques mucho mayores que los de caliza y arenisca, por lo que generalmente quedan *in situ* o no dejan entre ellos y el suelo un espacio apropiado para ser habitado; en cuanto a las pizarras, debido a lo riguroso del clima pirenaico, se desmenuzan casi totalmente originando superficies inhóspitas.

Los Carábidos altoaragoneses pueden dividirse en función de su hábitat en dos grandes grupos: los que habitan bajo piedras en áreas descubiertas y praderas y los que habitan entre las piedras mojadas de pequeño tamaño, que se encuentran en grandes cantidades en los bordes de los ríos y torrentes (graveras).

Las especies de ambos grupos se distribuyen, según sus preferencias, en los diferentes IP, dentro de un gradiente de variación de los mismos existente en el Alto Aragón. Destaca el hecho de que algunas especies que habitan en graveras consigan vivir alejadas de este hábitat siempre que sea en IP altos (mayor humedad) y que, por el contrario, algunas especies que habitan comúnmente en praderas vivan exclusivamente en graveras en IP bajos y medios (mayor humedad por proximidad al agua del río o torrente, lo que suple las deficiencias de humedad atmosférica por pluviosidad).

En principio, parece que la elección de hábitat obedecería a una característica fisiológica relacionada con la humedad y propia de cada especie.

En el Alto Aragón occidental ha habido épocas en que las altas presiones, al desbordarse por el occidente europeo, bloquearon el paso de los ciclones atlánticos y favorecieron unas condiciones climáticas continentales caracterizadas por veranos secos y cálidos y primaveras frías. Estos períodos alternaron con otros de tipo oceánico, caracterizados por veranos húmedos y frescos y primaveras suaves. El último intervalo de clima tipo continental se extendió desde el año 1934 hasta el año 1967. La fauna de Carábidos sufrió alteraciones en la última alternancia de períodos climáticos (años 60 y años 70), viéndose reducido el número de ejemplares de algunas especies endémicas (típicas de la zona) e incluso llegando a desaparecer totalmente algunas de ellas. Se descarta la posibilidad de que este hecho se deba a una influencia humana cada vez mayor. No obstante, y en un futuro, una recolección masiva de ejemplares, que por su belleza tientan a los coleccionistas, podría, junto con los mencionados cambios climáticos, contribuir a la desaparición de endemismos que actualmente son ya escasos.

La interacción entre los Carábidos y otros grupos de animales (en particular con otros depredadores que se cobijan bajo las piedras) se hace más patente

en los fondos de valle calizos en los que el agua se filtra con facilidad evitando el encharcamiento. Aquí, los animales que se alimentan de plantas serían devorados, en general, por Formícidos (hormigas) y Carábidos (también por otros animales), a excepción de los pulgones, que serían protegidos por algunas especies de hormigas. Los Carábidos depredan hormigas y las hormigas a su vez atacan (con más efectividad) a los carábidos. Los Arácnidos (arañas y opiliones), que son hostiles a Formícidos y Carábidos, serían depredados por estos últimos. Las musarañas, aves y anfibios depredarían tanto a los invertebrados carnívoros como a los consumidores de vegetales. Este equilibrio puede decantarse a favor de la acción de los Carábidos o de los Formícidos en áreas con diferente geomorfología. Así, en los fondos de valle con facilidad de encharcamiento desaparecerían la mayor parte de los Formícidos y en las partes bajas de las laderas soleadas el número de Carábidos (tanto de especies como de individuos) sería muy bajo, mientras que la presencia de Formícidos es patente, al contrario de lo que ocurre en las partes altas de estas mismas laderas.

En el Alto Aragón podemos considerar dos grandes grupos de especies de Carábidos, teniendo en cuenta su distribución espacial. Estaría integrado el primero por la mayoría de especies de Carábidos altoaragoneses, que viven generalmente en áreas descubiertas y praderas, bajo las piedras, pudiendo diferenciarse varios subgrupos característicos [*véase el esquema de la página 39*]:

1) Especies que colonizan casi uniformemente todo el hábitat. Podemos considerar como típica la asociación: *Calathus melanocephalus L.* y *Calathus fuscipes GOEZE.* Otras especies de características similares serían: *Ophomus rufipes DE GEER, Harpalus punctipennis MULSANT, Harpalus affinis SCHRANK, Steropus madidus FABRICIUS, Pterostichus dufouri DEJEAN, Calathus erratus C.R.SAHLBERG, Amara aenea DE GEER, Cyrtonotus aulicus PANZER.* En general, estas especies evitan los IP más bajos.

2) Especies que colonizan todo el hábitat, volando al sol en praderas o en lugares arenosos al borde de los ríos. La más característica es *Cicindela campestris L.*

3) Especies con tendencia a colonizar todo el hábitat, pero evitando siempre los IP más bajos y con gran preferencia por refugiarse en los bosques: *Procrustes purpurascens FABRICIUS, Chrysocarabus splendens OLIVIER*, ambas asociadas. Junto a éstas podemos considerar también a *Abax ater VILLERS.*

4) Especies con tendencia a colonizar todo el hábitat, pero evitando generalmente IP altos y los más bajos: *Peryphus hypocrita DEJEAN, Peryphus brunneicornis DEJEAN, Agonum mulleri HERBST, Anchomenus dorsalis PONTUPPIDAN.*

5) Especies que viven en IP altos. Las más representativas serían: *Cechenus pyrenaeus SERVILLE, Haptoderus sagittalis JEANNEL, Percosia equestris DUFTSCHMID, Pelor obesus SERVILLE.*

6)Especies que viven en índices de primavera bajos: *Brachynus ganglbaueri APFELBECK, Brachynus explode DUFTSCHMID.*

7) Especies que viven en IP altos y también en IP medios y bajos, pero en este último caso habitan exclusivamente graveras junto a arroyos y ríos: *Peryphus pyrenaeus DEJEAN, Princidium bipunctatum L.*, ambas asociadas. *Nebria lafresnayei SERVILLE, Poecilus lepidus LESKE, Cymindis humeralis FOURCROY.*

Tendríamos, por otro lado, otra serie de especies, que habitan entre las piedras mojadas de pequeño tamaño que se encuentran en grandes cantidades en los bordes de arroyos y ríos (graveras). Podemos diferenciar también varios subgrupos:

8) Especies más o menos extendidas en el hábitat aunque en índices de primavera altos habitan generalmente bajo piedras: *Peryphus decorus ZENKER, Peryphus testaceus DUFTSCHMID*, ambas asociadas. Estas especies que habitan graveras, pueden elegir entre seguir en graveras o habitar bajo piedras en prados en zonas con IP alto.

9) Especies que viven en zonas con IP medio y bajo: *Peryphus atrocoeruleus STEPHENS, Peryphus coeruleus SERVILLE,* ambas asociadas. *Agonum cyaneus DEJEAN.*

10) Especies que habitan zonas con IP bajo: *Synechostictus ruficornis STURM.*

En el Alto Aragón, con una superficie aproximada de 4500 kilómetros cuadrados, se han citado 195 especies de Carábidos, de las que 23 son endémicas del Pirineo. (A. Palanca Soler y C. Castán Lanaspa.)

Año de 10^{116}

Los cosmólogos han extrapolado las leyes de la física hacia atrás en el tiempo hasta una pequeña fracción de segundo tras la gigantesca explosión que originó nuestro universo. Menos suerte han tenido a la hora de extrapolar esas leyes hacia el futuro más remoto. Don N. Page y M. Randall McKee, de la Universidad estatal de Pennsylvania, acaban de describir cómo ven ellos el final del universo si continúa, indefinida, su expansión. Hablan de un cuadro evolutivo frío, difuso y caótico, donde el desorden aumenta sin cesar. El trabajo en cuestión lo han presentado en un número reciente de *Physical Review*.

Transcurrirán por lo menos 10^{116} años antes que llegue el universo de Page-McKee. Este futuro, inimaginablemente lejano, ha entrado en el terreno de la especulación gracias al desarrollo de teorías sobre la desintegración de partículas elementales. Para hacernos una idea de cuán remoto se halla todavía baste compararlo con la edad del universo: entre 8000 y 18.000 millones de años, es decir, unos 10^{10} años.

Page y McKee suponen que la densidad a gran escala de la masa en el universo es la misma en todas direcciones y lo suficientemente pequeña como para que la expansión continúe indefinidamente. Cuando haya cumplido su 10^{29} aniversario, el universo habrá expandido 10 billones de veces su volumen actual. Para entonces, las estrellas y galaxias habrán cedido a fuerzas gravitacionales internas que las harán colapsar hasta formar enanas negras, estrellas de neutrones o agujeros negros. El colapso puede absorber masas enormes, del orden de supercúmulos de galaxias, que crearán ingentes agujeros negros. Pero el modelo de Page y McKee (con una densidad global uniforme) supone que no se formarán agujeros negros con masa superior a un supercúmulo. Los conglomerados de masas existentes en esta fase, instados por fuerzas gravitacionales, pueden ser muy grandes, pero no sobrepasarán una masa total de 100 billones de veces la solar.

Cuando comience tan extraña era, la materia que haya sido atrapada por algún agujero negro se desintegrará. Argumentos teóricos, sujetos hoy a la corroboración experimental, sugieren que el límite inferior de la vida media del protón es de 10^{29} o 10^{30}. El neutrón, y demás partículas pesadas clasificadas como bariones, también se desintegrarán. Así pues, toda la materia que se encierre en los fríos restos de estrellas, y la que haya quedado desparramada como gas interestelar, se transformará en una sopa de electrones, positrones, neutrinos, fotones, gravitones y hasta cabe que otras partículas estables.

El electrón y el positrón son antipar-

tículas: se desintegran mutuamente cuando chocan. A una densidad relativamente alta, la de una enana negra o de una estrella de neutrones por ejemplo, abundarán semejantes colisiones. Los fotones de alta energía liberados en estas desintegraciones tenderán a redistribuir la energía sobre un ancho espectro térmico, que corresponderá a una temperatura de alrededor de un grado Kelvin en la superficie de una enana negra y de alrededor de 100 grados K. en la superficie de una estrella de neutrones. La emisión de fotones térmicos, con ser tan fría, seguirá dominando el fondo de radiación de microondas que, según se cree, al igual que el que existe en la actualidad, es un residuo de la explosión original. Ese tipo de radiación muestra hoy una temperatura de tres grados K., que en el año 10^{30} habrá descendido hasta los 10^{-13} grados K.

Los electrones y positrones resultantes de la desintegración bariónica tendrán mucha menos probabilidad de chocar en el gas interestelar extremadamente enrarecido. La distancia media entre dos partículas de éstas será del orden de una unidad astronómica, distancia que media entre el Sol y la Tierra. Mas, por llevar cargas opuestas, pueden formar un estado ligado si su energía cinética total es menor que su energía de enlace. El estado ligado se parece a un átomo de hidrógeno en el que el protón se ha sustituido por un positrón. Un tal átomo se llama positronio.

A medida que el universo continúa expandiéndose, las fuerzas atractivas entre partículas vecinas irán disminuyendo, pero la energía cinética de dichas partículas decrecerá algo más rápidamente. Imaginemos que el universo ha cumplido 10^{71} años: las fuerzas de atracción dominarán y empezará a formarse positronio interestelar. Por esa época el universo será 10^{40} veces mayor que ahora; los átomos de positronio serán estados de insospechado enrarecimiento de la materia que llenarán regiones del espacio tan grandes como el universo que podemos observar hoy.

En cuanto electrones y positrones empiecen a formar positronio, sólo será cuestión de tiempo el que se precipiten uno sobre el otro y se aniquilen. El tiempo requerido es 10^{45} veces mayor que la edad del universo en el momento del inicio de la formación. Vale decir que la aniquilación de gran parte de los electrones y positrones del gas interestelar durará por lo menos 10^{116} años. Las aniquilaciones producen fotones de alta energía, que quizá no lleguen a colisionar con la frecuencia exigida para que quedaran térmicamente distribuidos entre muchos estados energéticos.

Entretanto, los agujeros negros surgidos del colapso de estrellas y galaxias se irán aglutinando; mas ocurre también que hay efectos mecánico-cuánticos que permiten su evaporación. Un agujero negro de masa solar se evaporaría en 10^{66} años, mientras que un agujero negro resultante de un supercúmulo (alrededor de 10^{14} masas solares) lo haría en 10^{108} años. No se sabe cuál de entre los procesos, de aglutinación o evaporación, será el que prevalecerá, pero en todo caso la evaporación hacia el final de la historia del universo origina también fotones de alta energía que, con toda probabilidad, no se redistribuirán térmicamente.

Así, aunque la radiación de fondo del universo será fría en el año 10^{116} (de unos 10^{-70} grados K.), la desintegración de átomos de positronio y de agujeros negros nunca permitirá que la energía total del universo alcance una distribución térmica aleatoria. Según Page y McKee, el advenimiento de la muerte térmica nunca se producirá.

Estabilización del hidrógeno atómico

En situaciones normales, el gas se combina explosivamente para formar hidrógeno molecular. Una técnica, que inhibe esta reacción, permite investigar determinadas propiedades de un gas cuántico

Isaac F. Silvera y Jook Walraven

En el mundo de lo muy pequeño el dominio de la mecánica cuántica es absoluto. Pero, a escala macroscópica, los efectos mecánico-cuánticos son pocas veces relevantes y la sustancia que los muestre gozará de propiedades completamente distintas de las observadas en la materia ordinaria. Hasta la fecha, una de las sustancias cuánticas más sorprendentes ha sido el helio, que se convierte en un líquido cuántico a bajas temperaturas. En la fase de líquido cuántico, el helio fluye sin viscosidad y transporta el calor muchísimo mejor que los mejores conductores metálicos. No importa en cuánto disminuyamos su temperatura: no se solidificará a presión ambiente. Disponemos ya de otra sustancia dotada de propiedades cuánticas macroscópicas: un gas de hidrógeno atómico. En efecto, las propiedades del hidrógeno atómico a bajas temperaturas pueden resultar ser más extrañas que las del mismo helio. Así, el hidrógeno atómico se asemeja a éste en que no se solidifica, pero tampoco puede licuarse. Debido a que la teoría cuántica predice que permanece gaseoso a la temperatura del cero absoluto, se ha convenido en llamarle gas cuántico.

El hidrógeno constituye, por supuesto, un elemento familiar, pero no en su forma atómica. Un gas formado por átomos aislados de hidrógeno no es estable en las condiciones que prevalecen en la superficie de la tierra. El hidrógeno se encuentra fuertemente ligado a otros elementos para formar compuestos químicos, como agua, hidratos de carbono y otras sustancias orgánicas. Dos átomos de hidrógeno pueden también combinarse y producir una molécula diatómica estable (H_2). Aunque un gas de hidrógeno diatómico tiene propiedades cuánticas, si lo enfriamos suficientemente se licuará y después se solidificará. Desde hace mucho tiempo se sabe cómo disociar moléculas de hidrógeno diatómico en el laboratorio, pero los átomos resultantes se recombinan en menos de una milésima de segundo para formar nuevas moléculas.

A lo largo de los últimos años, junto con nuestros colegas de la Universidad de Amsterdam, hemos desarrollado un método para bloquear la recombinación del hidrógeno atómico, incluso a densidades relativamente altas. Hemos conseguido mantener un gas de átomos durante varias horas, a temperaturas extremadamente bajas y en presencia de un campo magnético intenso. En dicho estado, los átomos de hidrógeno se recombinan con bastante lentitud, lo que nos permite considerar el gas como estabilizado. Nuestros resultados experimentales demuestran que el hidrógeno atómico permanece en estado gaseoso hasta la temperatura más baja que hemos investigado, que es la de 0,08 grados Kelvin. Hallazgo que representa, cuando menos, una confirmación provisional de que el hidrógeno atómico se mantiene gaseoso en el cero absoluto, como predice la teoría cuántica.

Aunque constituye una sustancia nueva en el laboratorio, el gas de hidrógeno atómico estable no es algo inusual en el universo. El espacio interestelar está lleno de gas de hidrógeno atómico muy enrarecido, a pesar de la baja densidad reinante, que podríamos cifrar en un átomo por metro cúbico (comparado con las 3×10^{25} moléculas por metro cúbico del aire al nivel del mar), el gas interestelar constituye una gran parte de toda la materia del universo; y debe su carácter estable a la bajísima densidad en que se halla. La recombinación depende de las colisiones entre los átomos, tan raras que la vida media del gas es mayor que la vida esperada del universo.

En el laboratorio, sin embargo, se pueden producir muestras de hidrógeno atómico a temperaturas muy inferiores a las del espacio y con densidades de 10^{22} o 10^{23} átomos por metro cúbico, muchos órdenes de magnitud mayor que la densidad del gas interestelar. De acuerdo con los cálculos teóricos, tal densidad es sólo uno o dos órdenes de magnitud menor que la densidad necesaria para que se produzcan los extraños fenómenos cuánticos predichos para un gas cuántico, aunque todavía por observar.

El fenómeno cuántico más buscado es el de la condensación repentina de una gran proporción de los átomos del gas, en un estado de energía mínima. Se espera que la condensación tenga lugar a una temperatura baja, dependiente sólo de la densidad del gas. Por ejemplo, a una densidad de 10^{24} átomos por metro cúbico la temperatura crítica de condensación se sitúa en los 0,016 grados Kelvin; a la densidad del hidrógeno interestelar, sin embargo, la temperatura crítica desciende hasta 10^{-18} grados Kelvin. La temperatura crítica para la condensación es proporcional a la densidad elevada a la potencia 2/3.

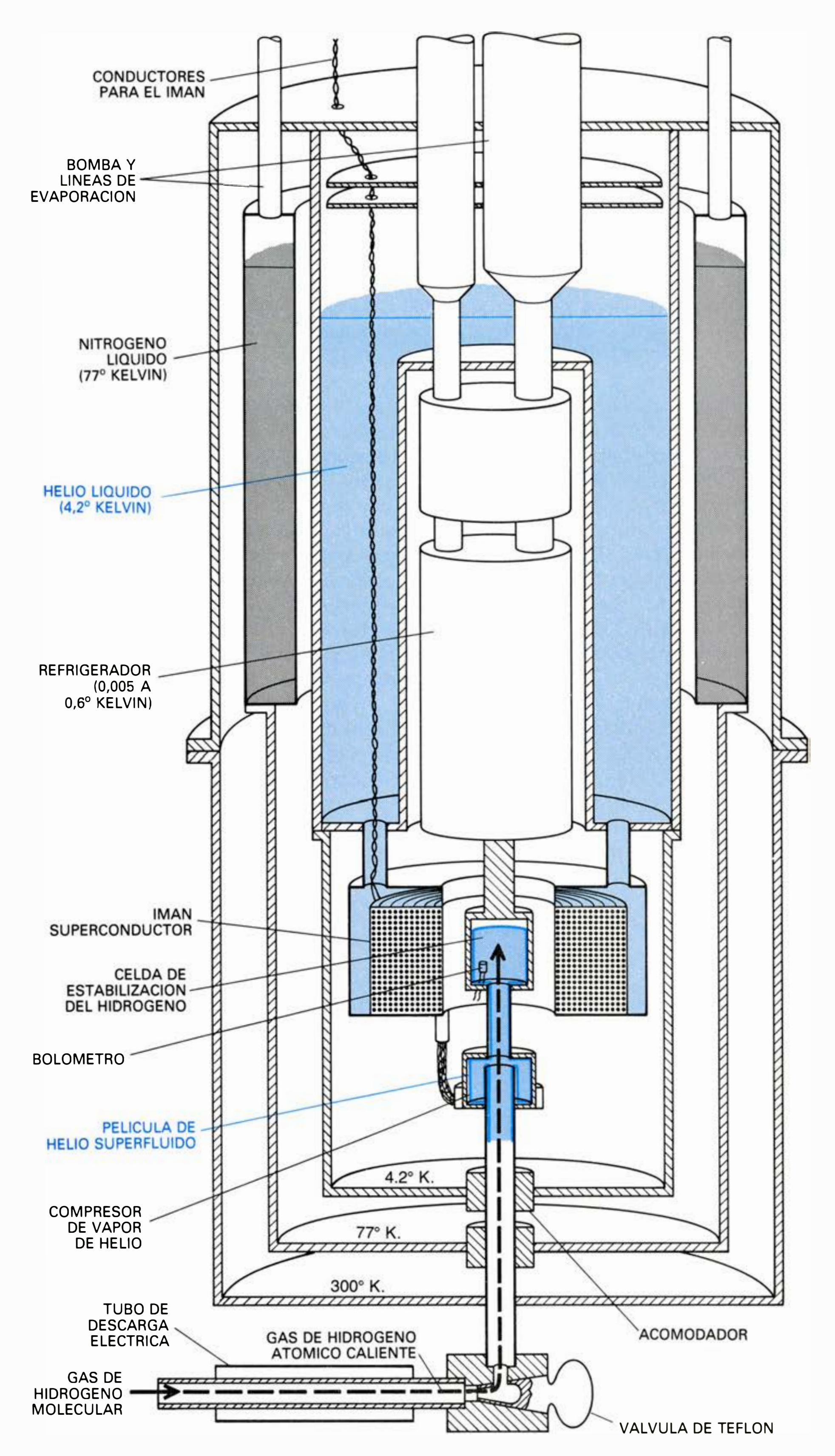
La idea de que un gran número de las partículas de un gas atómico puedan todas colocarse simultáneamente en su estado de energía más baja a temperatura no nula fue introducida por Albert Einstein en 1923. Según la termodinámica clásica del siglo XIX, o precuántica, sólo a la temperatura del cero absoluto una porción considerable de las partículas de un gas (de hecho todas ellas) pueden alcanzar sus niveles de menor energía. Por encima del cero absoluto, las partículas de un tal gas clásico se distribuirán entre muchos niveles

energéticos, ninguno de los cuales contendría una fracción importante de las partículas. La distribución de los estados de energía, para una temperatura dada, viene descrita por la estadística de Maxwell-Boltzmann, así llamada en honor de James Clerk Maxwell y de Ludwig Boltzmann. La estadística aporta la probabilidad de que un átomo tenga una cierta energía.

Sin embargo, en la descripción mecánico-cuántica correcta la distribución estadística de los estados energéticos, para un agregado de partículas idénticas, puede ser muy distinta. La teoría estadística que describe los átomos de hidrógeno fue estudiada por primera vez por el físico hindú S. N. Bose, por cuya razón se la denomina estadística de Bose. El fenómeno predicho por Einstein constituye una consecuencia matemática de la estadística de Bose, pero se oponía tanto a la intuición de los físicos de los años veinte que se la consideró pura rareza matemática, que nunca se encontraría en sistemas reales. Hoy, sin embargo, se cree que este fenómeno, denominado condensación de Bose-Einstein, es observable en el laboratorio.

Un gas de hidrógeno atómico, que se hubiera convertido en un condensado de Bose-Einstein, constaría de dos poblaciones de átomos: los condensados en el estado fundamental, o estado de menor energía, y los átomos que permanecieran en estados de energía más alta. Distribución de estados energéticos que quizá no nos parezca extraña si tenemos en cuenta que, en un gran agregado de átomos, hay probabilidad estadística de que algunos ocupen el estado fundamental. En la propia estadística de Bose, los átomos se hallan distribuidos entre numerosos estados energéticos a temperatura ordinaria; cada estado posee sólo una pequeña fracción del total. Mas en un gas condensado de Bose-Einstein, una buena porción de átomos ocuparían el estado fundamental a una temperatura alcanzable experimentalmente; y casi un 100 por ciento de los átomos se convertirían en átomos condensados a temperaturas bastante por encima del cero absoluto.

Suponía la termodinámica clásica que todo el movimiento atómico detuviese a temperatura cero; los átomos quedarían en reposo con energía cinética nula. En mecánica cuántica no cabe el estado sin movimiento, ya que violaría el principio de incertidumbre formulado por Werner Heisenberg. El principio establece que la posición y el momento de un átomo no pueden de-



APARATO EXPERIMENTAL desarrollado por los autores en la Universidad de Amsterdam: rompe las moléculas de hidrógeno y produce un gas de átomos de hidrógeno libres, que a continuación se estabiliza a bajas temperaturas en un campo magnético fuerte. El enlace químico que mantiene unidos los dos átomos en una molécula de hidrógeno se rompe fácilmente bombardeando la molécula con electrones energéticos; ahora bien, sin refrigeración y sin un campo magnético fuerte los átomos se recombinarían en menos de una milésima de segundo. El aparato incorpora varias unidades de refrigeración, unas dentro de las otras, mantenidas a temperaturas cada vez más bajas. El gas de hidrógeno atómico caliente es emitido por un tubo de descarga eléctrica y conducido a lo largo de una serie de superficies frías. El hidrógeno ya enfriado entra en un campo magnético, donde puede mantenerse varias horas. A una temperatura de 0,3 grados K se han alcanzado densidades de hasta 10^{23} átomos por metro cúbico durante unos minutos.

terminarse simultáneamente con precisión ilimitada. Si un átomo quedara totalmente en reposo podríamos medir su posición con exactitud. Se conocería también el momento de un átomo que no se moviera: por definirse el momento como el producto de la masa por la velocidad, el momento de un átomo estacionario debe ser cero.

El principio de incertidumbre implica que, aun cuando la temperatura de una sustancia pudiera reducirse al cero absoluto, los átomos retendrían una pequeña cantidad de energía cinética (la energía de punto cero). Sin embargo, el movimiento sería cualitativamente distinto del movimiento al azar característico de temperaturas más elevadas: todos los átomos tendrían la misma energía y, por tanto, el mismo estado de movimiento. Se espera que el movimiento coherente de los átomos condensados en un gas condensado de Bose-Einstein dé origen a propiedades macroscópicas extraordinarias a temperaturas muy por encima del cero absoluto.

Los átomos que constituyen alrededor del 50 por ciento de los elementos están sujetos a la estadística de Bose. Vale la pena, pues, pararse a pensar por qué la condensación de Bose-Einstein no se ha convertido en un fenómeno común, observado en muchas sustancias a bajas temperaturas. La razón es que la condensación toma su forma más simple (y la forma en que la describió Einstein) sólo en un gas ideal, aquel que está compuesto de átomos que no interaccionan entre sí. Todos los átomos reales tienen interacciones electromagnéticas que tienden a ligar unos con otros; en virtud de cuyas interacciones la mayoría de las sustancias se solidifican mucho antes de alcanzar la temperatura a la que podría tener lugar la condensación de Bose-Einstein.

Para que la condensación llegue a observarse, los movimientos de punto cero habrán de ser lo suficientemente importantes y dominar así sobre la atracción electromagnética de los átomos. La magnitud del movimiento de punto cero aumenta a medida que se reduce la masa atómica o la intensidad de las interacciones electromagnéticas. Por tanto, los átomos químicamente inertes de los elementos más ligeros son los únicos candidatos plausibles. El helio tiene propiedades cuánticas interesantes porque es, a la vez, ligero e inerte, de modo que las fuerzas interatómicas son débiles. Por esas mismas razones, el helio no se solidifica.

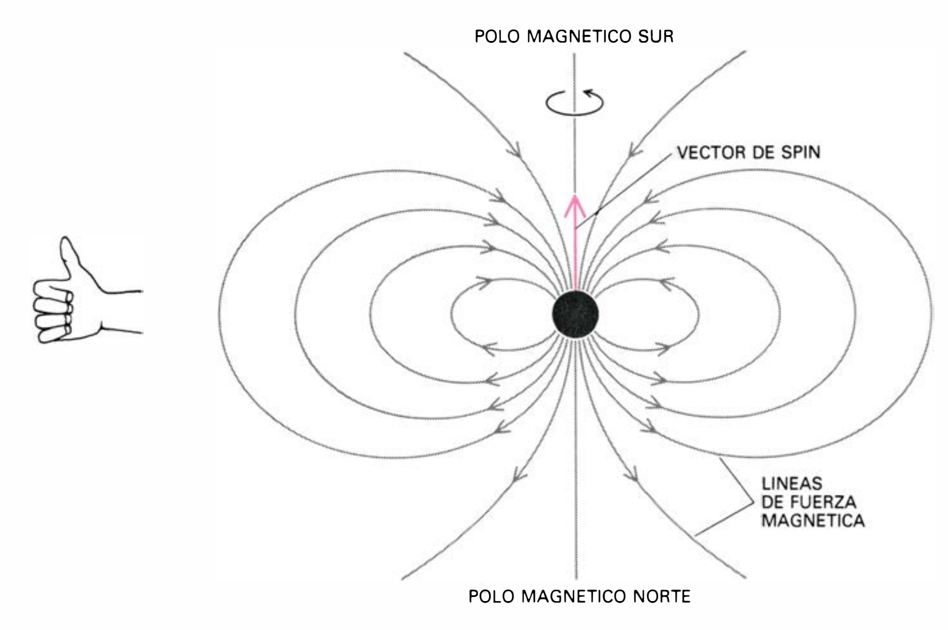
Ni que repetirse tiene que el hidrógeno es más ligero y posee movimientos de punto cero más importantes que los de cualquier otro átomo. Sin embargo, bajo condiciones ordinarias los átomos de hidrógeno están sujetos a fuerzas interatómicas fuertes y forman moléculas diatómicas. Sometiendo los átomos a condiciones extremas, puede lograrse que las fuerzas interatómicas sean más débiles que las que gobiernan cualquier otro sistema atómico. En esas circunstancias excepcionales, basta la energía cinética de los átomos, incluso en el estado de energía más baja, para evitar que el hidrógeno se solidifique o se licúe. La sustancia resultante quizá represente la aproximación más cercana posible al gas cuántico ideal imaginado por Einstein y ofrezca la mejor oportunidad para observar la condensación de Bose-Einstein.

Cuanto se puede esperar del hidrógeno atómico estable se lo debemos a 70 años de investigaciones de las propiedades del helio a bajas temperaturas, en particular de las propiedades del isótopo común helio 4. (El núcleo de un átomo de helio 4 consta de cuatro partículas: dos protones y dos neutrones. Otros isótopos del helio se distinguen en razón del número de neutrones en el núcleo atómico.) El helio 4 se convierte en un líquido ordinario a 4,2 grados Kelvin. Si con una bomba de vacío aligeramos la presión de vapor reinante sobre el helio, el líquido hervirá violentamente y se enfriará más, por consiguiente. La formación de burbujas en un líquido en ebullición está causada por conductividad térmica relativamente baja.

¿Qué ocurre cuando disminuye la temperatura del helio hasta 2,18 grados? El líquido se transforma de repente en tan buen conductor del calor que cesa la ebullición. Evaporación y enfriamiento continúan en la superficie, pero sólo las corrientes de convección redistribuyen la energía térmica remanente a través del volumen. Tales corrientes pueden suministrar transporte adecuado de calor en un fluido que se mueve sin viscosidad, o resistencia a fluir.

Así, se dice que el helio 4 líquido a 2,18 grados o por debajo de esta temperatura es un superfluido. Si se le hace fluir en un tubo cerrado, sobre sí mismo, el líquido prosigue en su avance sin fricción y nunca se llega a parar, como lo haría un fluido normal. Fluye en los conductos más pequeños de la vasija que lo contiene, y tiene la notable propiedad de discurrir a través de polvo densamente apretado, haciendo caso omiso de la barrera. Una vasija con agujeros microscópicos, que serían impenetrables para un fluido normal, puede filtrar el superfluido, como si fuera ella un tamiz. Diremos que una vasija así tiene una superfuga.

Pero incluso sin tal superfuga, la vasija abierta y llena con un superfluido no tardará mucho tiempo en ir perdien-

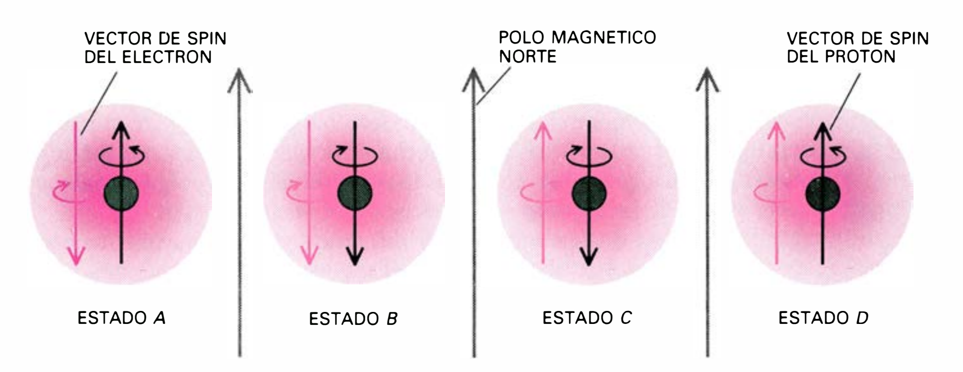


MOMENTO ANGULAR DE SPIN es una propiedad de las partículas que componen un átomo de hidrógeno, que puede ser manipulada para evitar que los átomos se recombinen. El spin es una magnitud vectorial, dotado, por tanto, de cantidad y dirección. La magnitud del vector puede interpretarse como el momento angular de rotación, mientras que la dirección del vector da el sentido de la rotación de acuerdo con la regla de la mano derecha. (Si los dedos se curvan en el mismo sentido que la rotación, el pulgar indica la dirección del vector.) Una partícula que posea masa y spin generará un campo magnético; la orientación de la partícula se hallará sujeta a la acción de un campo magnético externo aplicado. Para una partícula de carga negativa, el polo norte magnético apunta en la dirección opuesta al vector spin.

do contenido. Una película delgada del superfluido, de unos 200 angstrom de grosor, empezará a trepar por las paredes interiores del recipiente, alcanzará el borde y descenderá por las paredes exteriores. Este fenómeno es otra consecuencia de la falta de viscosidad. Todos los fluidos que pueden mojar una superficie tienden a trepar por ella en virtud de las fuerzas adhesivas entre los átomos del fluido y los átomos de la superficie. En principio, cualquier líquido debería trepar por las paredes del recipiente que lo contiene y reducir al mínimo su energía total, descendiendo a un nivel más bajo. En fluidos ordinarios, empero, el flujo de la película lo inhibe la viscosidad. Por tanto, un fluido sin viscosidad, así el helio superfluido, forma una película continua entre el interior y las paredes exteriores más bajas de la vasija, terminando ésta por vaciarse.

Hay propiedades no menos sorprendentes de los superfluidos que no pueden explicarse por la mera falta de viscosidad. Pensemos en la tendencia a moverse hacia fuentes caloríficas; en una película delgada, el flujo llega a alcanzar los 50 centímetros por segundo. Si el calor se aplica al extremo superior de un tubo delgado, cuyo extremo inferior esté sumergido en helio 4 superfluido, la película de helio irá tubo arriba y se evaporará en la fuente caliente. Si el extremo inferior del tubo se llena con un polvo fino (que un fluido normal no pudiera penetrar ni siquiera bajo presión), el superfluido avanzará hacia arriba con tal rapidez que originará un surtidor en el extremo superior del tubo y caerá de nuevo en la vasija.

No está definitivamente establecido por vía experimental, pero resulta muy probable que el helio 4 superfluido sea un condensado de Bose-Einstein. Para confirmar la hipótesis habría que medir la fracción de los átomos en el estado fundamental, lo que hasta la fecha no se ha podido conseguir de una forma inequívoca. Fritz London, de la Universidad de Duke, propuso por primera vez esta hipótesis, en 1938, cuando demostró que algunas de las propiedades observadas del helio superfluido admitían explicación en la suposición de que se tratara de un condensado de Bose-Einstein. Los átomos de helio 4 están sujetos a la estadística de Bose. Por lo que London propuso que un superfluido está formado por dos componentes: un condensado, o componente superfluida, que puede atravesar pequeños intersticios, y una componente normal, que no puede.

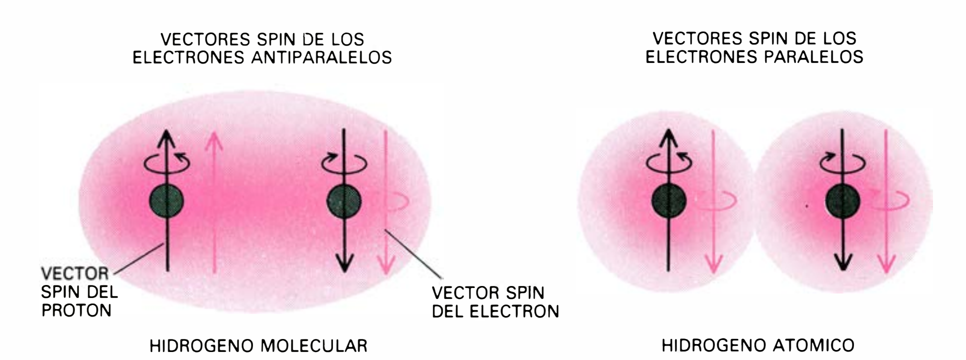


SPIN DEL ATOMO DE HIDROGENO. No es más que la suma de los vectores de spin del protón y del electrón constituyentes. Tanto el vector de spin del electrón como el del protón están cuantizados: la magnitud de ambos vectores de spin vale 1/2 (cuando se mide el momento angular en unidades fundamentales) y los vectores pueden apuntar hacia arriba o hacia abajo. El hidrógeno podrá adoptar sólo cuatro configuraciones en relación con su spin. Los estados mecánico-cuánticos de spin resultan idénticos a las cuatro configuraciones de un campo magnético infinito. La densidad de la nube electrónica que rodea cada núcleo atómico representa la probabilidad relativa de encontrar un electrón en una región dada.

Así, en el efecto surtidor sólo el condensado superfluido puede penetrar a través del polvo y alcanzar la parte superior del tubo. El efecto de calentar excita los átomos hasta estados de energía más alta, convirtiendo la componente superfluida en un fluido normal. Para mantener una distribución uniforme de la componente superfluida a través del volumen del fluido, debe seguir subiendo por el tubo una fracción de dicha componente.

No obstante esos éxitos cualitativos, el modelo de London ofrece serios inconvenientes que han sido objeto de análisis teóricos durante décadas. Mientras que la predicción de Einstein se fundaba en un gas de partículas que no interaccionan entre sí, los efectos de las fuerzas interatómicas de un líquido son importantes y no pueden despreciarse. Los átomos en el estado fundamental de tal gas tienen momento nulo, por lo que la incertidumbre sobre su posición es máxima. Las interacciones modifican el estado fundamental de los átomos en un líquido, de ahí que pocos átomos del estado fundamental posean momento nulo. Si un líquido ha de pasar entonces a convertirse en un condensado de Bose-Einstein, los átomos condensados deberán ocupar un estado fundamental más complejo que el estado fundamental de los átomos de un gas. Por si fuera poco, resulta muy difícil identificar, experimentalmente, el estado fundamental de un líquido.

Todos esos problemas pueden evitarse, en gran parte, cuando se trata de un gas superfluido de hidrógeno atómico. Sabido es que los líquidos gozan de una densidad fija a grandes rasgos. No así los gases; la de éstos podemos modificarla controlando la presión en la vasija que los contiene. En un gas a baja densidad, los efectos de las interacciones atómicas son mucho menos impor-



ESTABILIZAR EL HIDROGENO ATOMICO exige que los vectores de spin de los electrones estén polarizados u orientados en la misma dirección. Dos átomos cuyos vectores spin de los electrones estén en direcciones opuestas tenderán a compartir sus electrones y se ligarán fuertemente para formar una molécula diatómica. En ausencia de fuerzas polarizadoras fuertes, los vectores de spin de los electrones están orientados al azar, y no tardará mucho en que los átomos de hidrógeno constituyan pares. Pero si los vectores spin de los electrones son paralelos, no pueden compartirse los electrones, las fuerzas interatómicas son débiles y los átomos de hidrógeno permanecen en estado monoatómico. Sólo hay que prestar atención a las orientaciones relativas de los dos vectores spin de los electrones (si son paralelos o antiparalelos) para entender la formación de hidrógeno diatómico; las direcciones reales de los vectores, así como el spin de los protones, poseen menos importancia. (Aparecen dos de los 16 pares posibles de átomos.)

tantes que en un líquido. Cálculos teóricos sugieren que casi el cien por ciento de los átomos condensados en un gas estarán en un estado de momento nulo. Más aún, las propiedades, magnéticas y otras, del átomo de hidrógeno que no comparte con el átomo de helio prometen enriquecer la variedad de fenómenos macroscópicos que pueden observarse en un superfluido.

No es difícil conseguir hidrógeno atómico. Mas su estabilización requiere resolver varias dificultades experimentales en una área donde se carecía del entrenamiento suficiente. Aun cuando nos proponíamos examinar en el laboratorio ciertas predicciones de la teoría cuántica, el diseño de nuestro aparato experimental no habría sido posible sin la guía de otras predicciones de la teoría.

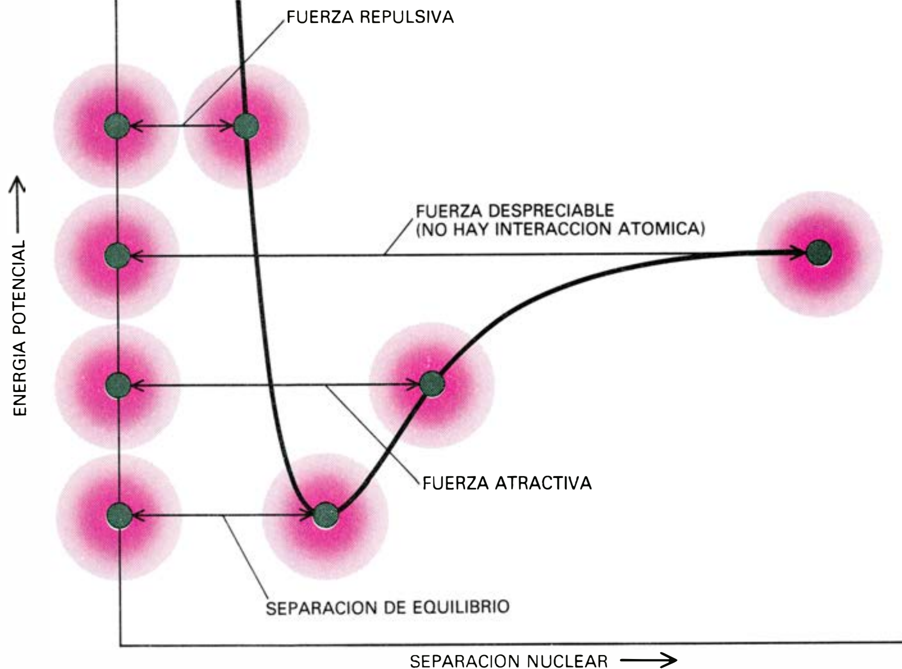
Una de las propiedades cuánticas más importantes de un átomo o de una partícula subatómica es su spin, o momento angular intrínseco. El spin de una partícula es similar al giro de una bola. Se trata de una cantidad vectorial, dotada, por tanto, de magnitud y dirección. La magnitud del vector de spin puede interpretarse como una medida de la velocidad angular de rotación; la dirección del vector especifica el sentido de la rotación, de acuerdo con la convención llamada regla de la mano derecha. Si los dedos de la mano derecha se doblan en el mismo sentido que la rotación, el dedo pulgar señala la dirección del vector de spin.

A diferencia de la rotación de una bola, el spin de un átomo o de una partícula subatómica está cuantificado. La magnitud del spin sólo puede tomar valores enteros o semienteros (cuando el momento angular se mide en unidades fundamentales) y el vector dispondrá de un número finito de direcciones. El protón, el neutrón y el electrón, por ejemplo, tienen un spin de magnitud 1/2; el vector de spin sólo tomará dos direcciones: "arriba" y "abajo". La partícula cuyo spin posee valores semienteros, tales como 1/2, 3/2, 5/2 y así sucesivamente, se llama fermión, en honor de Enrico Fermi.

Las partículas cuyo spin tiene una magnitud que es un entero de la serie 0, 1, 2, 3,... se denominan bosones, en honor de Bose. El fotón, o cuanto de la radiación electromagnética, es un bosón; también lo es el pion, el más importante mediador de las fuerzas de ligadura en un núcleo atómico. Un bosón con spin 0 es una partícula que, claro está, carece de spin; el vector de spin no será necesario entonces para describir sus propiedades. Para un bosón con spin 1, el vector tiene tres direcciones posibles.

El spin de una entidad compuesta, piénsese en el hidrógeno atómico, depende de la intensidad del campo magnético donde el átomo se encuentra inmerso. El caso de explicación más sencilla corresponde a un campo magnético infinito; para un campo magnético fuerte, un campo infinito constituye una buena aproximación. En un campo infinito, el spin del hidrógeno atómico es el vector suma de los vectores spin de las partículas que lo forman. En general, la suma depende de la orientación del spin de los constituyentes. Un átomo de hidrógeno, que consta de un protón y un electrón, puede existir en uno cualquiera de cuatro estados de spin. Los vectores del spin del protón como del electrón pueden estar dirigidos hacia arriba, de forma que el spin del átomo es 1 en la dirección hacia arriba, o ambos vectores spin de los constituyentes pueden estar dirigidos hacia abajo, de forma que el spin atómico es 1 en la dirección hacia abajo. En las configuraciones restantes, los vectores de spin de los constituyentes están dirigidos en sentidos opuestos, el protón hacia arriba y el electrón hacia abajo, o el protón hacia abajo y el electrón hacia arriba, de suerte que el spin total del átomo en cuestión será cero. En todas las orientaciones posibles resulta manifiesto que el átomo de hidrógeno es un bosón.

La distinción entre fermiones y bosones aportó la clave de la técnica que ha permitido estabilizar el hidrógeno atómico. Poseen los fermiones una propiedad característica: obedecen una regla mecánico-cuántica llamada el principio de exclusión, formulado por Wolfgang Pauli. Si un fermión ocupa cierto estado, los demás fermiones idénticos se hallan excluidos de ese estado. En mecánica cuántica, el estado de una partícula nos define, a modo de completa especificación, su identidad y circunstancias en que se halla, incluyendo su energía, posición o momento y spin. Si dos electrones ocupan regiones que se superponen en el espacio y tienen idéntica energía y momento, diferirán en su spin. Los bosones no obedecen, por contra, el principio de exclusión; no hay límite en el número de bosones que pueden ocupar un solo estado. En efecto, en la condensación de Bose-Einstein todos los bosones de una



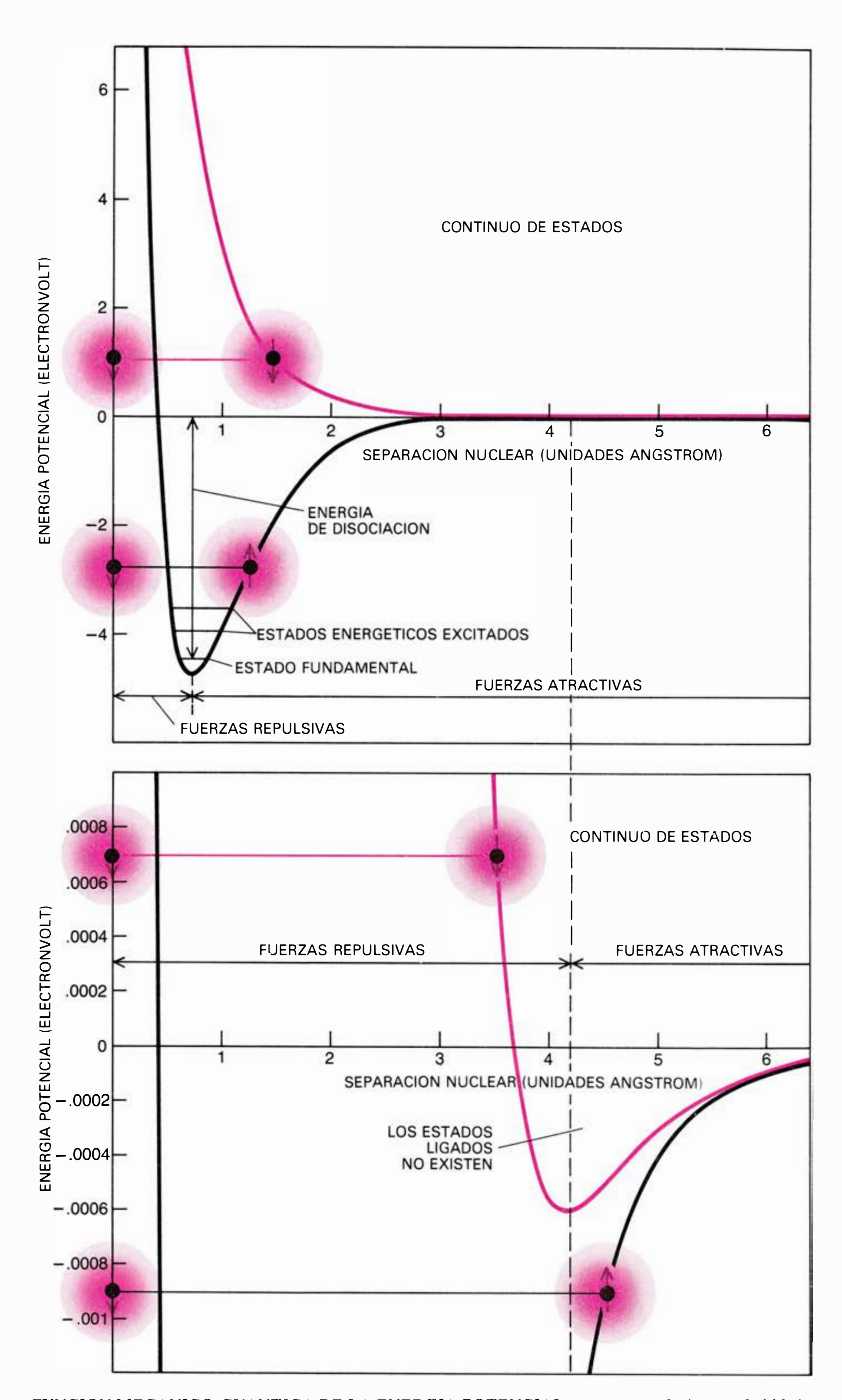
ENERGIA POTENCIAL de dos átomos de hidrógeno en función de su distancia mutua. Para una separación muy grande, la energía potencial se iguala, arbitrariamente, a cero. En cada punto de la curva, salvo en el mínimo, hay una fuerza neta sobre los átomos que tiende a llevarlos hasta una energía más baja. El punto de energía potencial mínima representa la separación de equilibrio o posición de reposo del segundo átomo con respecto al primero. Si el segundo átomo se acerca al primero, la energía potencial de los átomos aumenta y hace acto de presencia una fuerza repulsiva neta entre ellos que insta a restaurar la posición de equilibrio. Del mismo modo, si se aleja el segundo átomo del primero, la energía potencial aumenta de nuevo y una fuerza atractiva tiende a situar los átomos en la posición de equilibrio. Si los átomos se alejan suficientemente, un pequeño cambio en su separación no altera mucho la energía potencial del sistema; una pequeña cantidad de energía cinética les permite incluso moverse libremente sin interacción importante. En esta función hipotética de la energía potencial, no se ha tenido en cuenta las condiciones impuestas por la mecánica cuántica sobre la energía de los átomos y sobre sus interacciones.

muestra macroscópica de materia pueden, en principio, converger al mismo estado, el estado fundamental.

Un átomo de hidrógeno aislado puede tratarse como un bosón compuesto, pero no dos átomos que se aproximen el uno al otro. En la formación de una molécula de hidrógeno diatómico, hay que prestar atención a las propiedades de los dos electrones y de los dos protones que la integran. Cuando los dos átomos se aproximan, sus nubes de electrones periféricos empiezan a superponerse. Dos electrones cuyos vectores de spin estén en la misma dirección no pueden ocupar la misma región del espacio simultáneamente. Pero si los vectores spin de los electrones apuntan en direcciones opuestas, ningún principio cuántico les impedirá que estén en el mismo lugar al mismo tiempo. (La repulsión electrostática tiende a apartar los dos electrones, pero esto es aquí una consideración secundaria.)

Dos átomos de hidrógeno próximos tienden a adquirir la configuración de menor energía posible. Al hacer mínima la energía, electrones y protones constituyentes de los átomos tienden a ordenarse de modo que las fuerzas electrostáticas atractivas y repulsivas que actúan entre ellos queden compensadas exactamente. Para un sistema de dos átomos de hidrógeno, la configuración de energía más baja depende, casi exclusivamente, del estado de los electrones. Por ejemplo, si los vectores de spin de los dos electrones son antiparalelos, las cargas negativas de los electrones pueden interponerse entre los núcleos positivamente cargados y compensarán en parte la repulsión nuclear. Con esta configuración de los electrones, la energía total del sistema es mucho menor cuando los átomos están juntos que cuando se hallan distantes. Los electrones constituyen un fuerte enlace cohesivo. Excepto a temperaturas altas, los átomos tienen una energía cinética térmica insuficiente para romper el enlace; los átomos han formado una molécula diatómica de hidrógeno que es estable.

Si el spin de los electrones es paralelo, no se puede concentrar entonces la carga electrónica entre los núcleos sin violar el principio de exclusión. Las fuerzas resultantes entre los dos átomos son repulsivas, excepto para una separación nuclear de alrededor de cuatro angstrom, pues aparece entonces una pequeña fuerza atractiva. Sin embargo, la fuerza atractiva es muy débil, hasta el extremo de que basta la energía de



FUNCION MECANICO-CUANTICA DE LA ENERGIA POTENCIAL para un par de átomos de hidrógeno con los vectores de spin de los electrones antiparalelos (*negro*), que es totalmente distinta de la función para un par de átomos con los vectores de spin paralelos (*color*). Los átomos tienden a adoptar una configuración que minimice su energía potencial, pero no pueden permanecer en reposo a la separación de equilibrio, ya que en mecánica cuántica la posición de una partícula no puede ser fijada con toda precisión. Lo que sucede, si los átomos están ligados, es que deben ocupar uno de una serie de estados energéticos; el más bajo, o estado fundamental, los mantiene lo más cerca posible de la posición de equilibrio. Si los átomos no están ligados, pueden tomar cualquier estado de energía positiva con un continuo de valores. Cuando los dos vectores de spin electrónicos son antiparalelos, el estado fundamental se halla cerca del fondo de un pozo de potencial profundo y la energía necesaria para disociar los átomos es grande. Por tanto, los átomos minimizan su energía total formando un enlace molecular. El enlace, suficientemente fuerte, resiste los movimientos térmicos de los átomos hasta temperaturas muy altas. Sin embargo, cuando los vectores de spin son paralelos las fuerzas atractivas son muy débiles, tanto que incluso los débiles movimientos que la mecánica cuántica predice en el cero absoluto impedirán que los átomos formen un enlace. Por tanto, no existe estado fundamental. En el diagrama inferior de la ilustración, la escala vertical está aumentada en un factor de unos 5000 con relación a la del diagrama superior.

punto cero de los átomos para contrarrestar la fuerza y evitar la formación de una molécula. La razón fundamental de que un gas de hidrógeno atómico no sea estable bajo las condiciones usuales que hay en la superficie de la tierra reside en que los vectores de spin de los electrones muestran, por así decirlo, una distribución al azar. Casi siempre podrá un átomo de hidrógeno aparearse con otro átomo de hidrógeno cuyo vector de spin del electrón apunte en la dirección opuesta.

Nuestra estrategia fundamental para crear hidrógeno atómico estable consistió en polarizar los vectores spin de los electrones, de forma que todos ellos quedaran orientados en la misma dirección. Cuando la polarización es total, los átomos no pueden recombinarse. La sustancia que se obtiene se denomina hidrógeno atómico con spin polarizado.

¿Cómo conseguir hidrógeno con spin polarizado a partir de hidrógeno atómico ordinario? Las partículas dotadas de masa y spin tienen un vector momento magnético asociado con su spin. Por tanto, el estado de spin de un electrón puede mantenerse mediante un campo magnético externo, exactamente igual que el campo magnético terrestre mantiene la dirección de la aguja de una brújula. El grado en que un conjunto de electrones atómicos puede polarizarse se representa como el cociente del número de átomos que tienen los electrones con spin hacia arriba (átomos con spin hacia arriba) al número de átomos que tienen los electrones con spin hacia abajo (átomos con spin hacia abajo). El grado de polarización alcanzado guarda una dependencia estrecha con la intensidad del campo magnético y la temperatura. En nuestros experimentos, el cociente de ambas orientaciones del spin es de 1 en 10^{20}. Se logra mediante un campo magnético de unos 100.000 gauss (200.000 veces más fuerte que el campo magnético de la Tierra) y a 0,3 grados K. de temperatura.

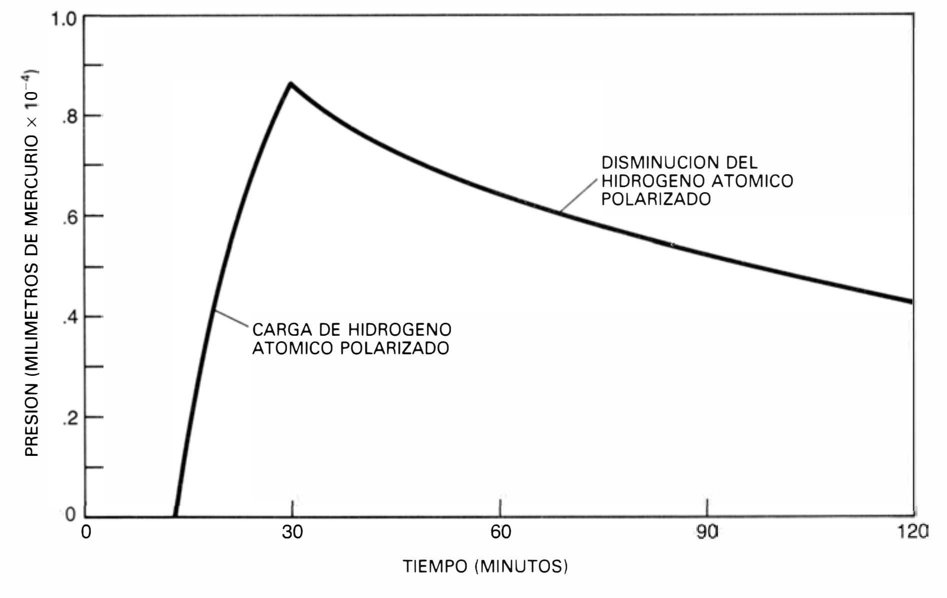
Pero a diferencia de la aguja de una brújula, cuya orientación puede cambiarse o mantenerse con un campo magnético, el estado de spin de un electrón no puede cambiarse mediante un campo magnético uniforme, por muy fuerte que sea éste. Para polarizar los electrones, necesitamos usar el efecto de un campo magnético de gradiente intenso: un campo que varía rápidamente de un lugar a otro. El campo de gradiente intenso no es más eficaz que otro uniforme para invertir el sentido de los spin, pero el gradiente puede fácilmente separar los átomos con spin hacia arriba de los átomos con spin hacia abajo.

Los átomos con spin hacia abajo se ven forzados hacia las regiones que poseen campos magnéticos más intensos, donde se mantienen en sus estados de polarización durante varias horas. (En nuestro aparato, los campos magnéticos más intensos coinciden con las regiones de temperatura más baja.) Los átomos con spin hacia arriba son repelidos hacia las regiones dotadas de campos magnéticos más débiles, y chocan con las paredes que limitan el aparato. Allí, pasan a tener spin hacia abajo, yéndose hacia las regiones de campo intenso, o bien se recombinan, formando hidrógeno diatómico. El hidrógeno diatómico queda ligado de una manera permanente a las paredes frías, sin mayor incidencia posterior.

Los átomos con spin hacia abajo entran en una celda de estabilización al final de un tubo, donde el gradiente del campo magnético y la baja temperatura sirven para confinarlos. Puesto que los átomos polarizados viajan en dirección contraria al gradiente del campo magnético, se aceleran. Si las paredes de la celda de estabilización no estuvieran enfriadas, el aumento neto de la energía cinética de los átomos que entran en la celda bastaría para permitirles avanzar contra el gradiente de campo magnético y escapar así de la celda. Cuando los átomos llegan a las paredes frías de la celda pierden allí su energía cinética, quedando atrapados por la barrera magnética.

Una vez que hubimos entendido cómo separar los átomos de hidrógeno y cómo atraparlos en un estado de spin polarizado, gran parte de nuestro esfuerzo empleado en el diseño del aparato experimental tuvimos que dedicarlo a evitar que los átomos se pusieran sobre las paredes del aparato y se recombinaran allí. En hidrógeno con spin polarizado, un choque entre dos átomos puede conducir, a veces, a la inversión del spin de uno de los átomos y a su recombinación molecular subsiguiente. La colisión debe venir mediada por una tercera partícula para que se conserve la energía total y el momento del sistema. Cuando dos átomos se enfrentan en un choque aislado, sin que intervenga una tercera partícula que se lleve la energía liberada en la colisión, la inversión del spin está prohibida por las leyes de conservación. De donde se sigue que, en un gas, habrán de chocar simultáneamente tres partículas, por lo menos, para que ocurra la recombinación; a baja densidad, la probabilidad de un suceso de este tipo será, además, pequeña.

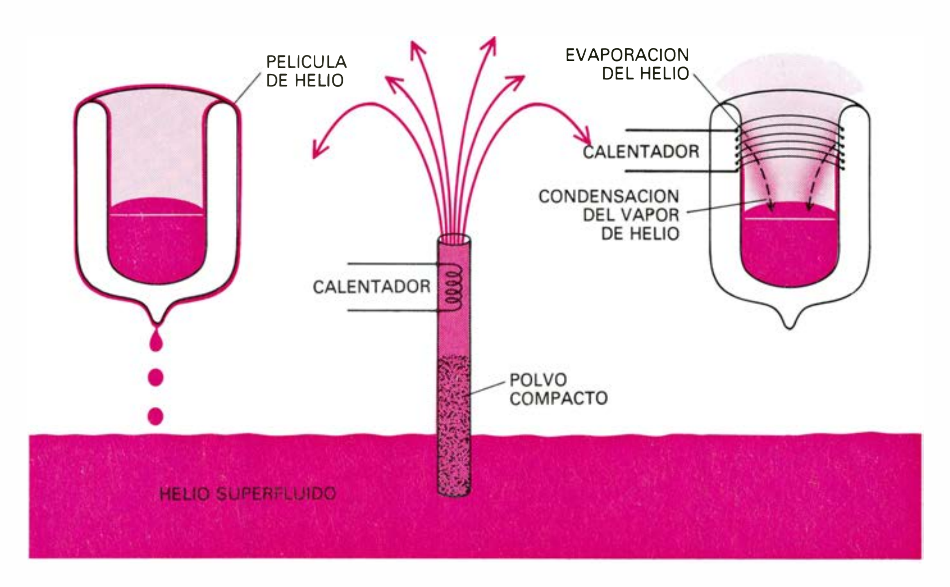
En la superficie, sin embargo, cualquier colisión entre dos átomos puede terminar en recombinación, pues los átomos de la misma pueden servir de catalizadores, como un tercer cuerpo. El problema se complica notablemente ante la necesidad de que las paredes del aparato se mantengan a temperatura baja para enfriar así el gas hidrógeno. De no recubrir las paredes con materiales especiales, las fuerzas electromagnéticas entre los átomos de hidrógeno y



AUMENTO DE LA PRESION del gas de hidrógeno estabilizado con spin hacia abajo durante la carga en la celda de estabilización; viene seguido por una disminución gradual tras la detención de la carga. La disminución es causada por la lenta recombinación de los átomos para formar moléculas diatómicas. De acuerdo con la teoría cuántica, en un campo magnético de intensidad finita una fracción de los átomos polarizados se halla en un estado de mezcla, con una pequeña probabilidad de que un electrón esté tanto hacia arriba como hacia abajo. La presencia de los estados mezcla abre una puerta a la recombinación.

aquéllas se volverán tan intensas a bajas temperaturas que ligarán los átomos a las paredes. Conforme aumenta la densidad de los átomos adheridos a las paredes, crece el ritmo de recombinación; sépase que dicho valor varía como el cuadrado de la densidad superficial. La eficacia de la recombinación superficial puede llegar hasta el punto de que el propio hidrógeno atómico de spin polarizado se recombine totalmente, en menos de una milésima de segundo, sobre superficies ordinarias.

Para evitar la recombinación superficial, hay que dar con una superficie que posea fuerzas atractivas débiles con el hidrógeno atómico. La mejor para este propósito sería el propio hidrógeno atómico con spin polarizado; pero si uno pudiera mantener tal superficie, la adhesión superficial habría dejado de ser un problema. Un caso que muestra afinidad débil para el hidrógeno e inhiba su acumulación lo tenemos en la superficie de helio líquido. Constituye ésta una película delgada que cubre todas las superficies de la celda de estabilización; a las bajas temperaturas necesarias para el experimento, se torna superfluida. Hay que atravesar secuencias de enfriamiento especiales para emplear una película superfluida y mantener todavía la baja temperatura de la celda.



EL HELIO SUPERFLUIDO interviene en varios fenómenos característicos de un líquido cuántico; algunos de estos fenómenos pueden tener sus análogos en un gas de hidrógeno atómico, y algunos de ellos han sido usados en la preparación del gas. Si el superfluido se coloca en una vasija abierta, una película trepará por las paredes laterales, pasará por el borde y descenderá por la parte exterior. Pasará a través de un polvo muy fino y compacto que sería impenetrable a un fluido normal, y se moverá a través de los estrechos poros del polvo a la velocidad de 50 centímetros por segundo hacia una fuente térmica. Si el extremo superior de un tubo vertical con polvo prensado se calienta, el helio superfluido subirá tan rápidamente por el tubo que formará una fuente. El calor causará que una película de fluido trepe por los lados de un tubo, donde se evaporará, volviendo a la región más fría que está encima del líquido y allí se condensará. Se usó la circulación del vapor de helio para facilitar la compresión del hidrógeno atómico.

Se comienza por producir hidrógeno atómico, a partir de gas hidrógeno molecular, en un tubo de descarga eléctrica. En el tubo, un campo eléctrico acelera los electrones libres, que chocan con las moléculas de hidrógeno diatómico, disociándolas. El bombardeo, muy intenso, de electrones determina que el gas atómico salga del tubo a una temperatura de varios centenares de grados Kelvin.

El gas caliente corre, con rapidez, hacia el interior frío de la celda de estabilización, a través de un tubo cuyo interior se halla recubierto con el polímero de fluorocarbono denominado teflón. Resulta curioso que el mismo material que forma un recubrimiento antiadhesivo para sartenes sirva también para evitar que el hidrógeno atómico se pegue a las paredes de un tubo en las primeras secuencias del enfriamiento. En su movimiento por el tubo, los átomos de hidrógeno chocan repetidas veces contra las paredes de teflón refrigeradas, perdiendo en cada ocasión algo de su energía cinética. Así se enfría el gas. Suelen bastar varios choques por átomo para que el gas alcance el equilibrio térmico con la pared. Si la temperatura de la pared de teflón supera los 20 o 30 grados Kelvin, y el paso del gas es breve, no habrá apenas recombinación de los átomos de hidrógeno.

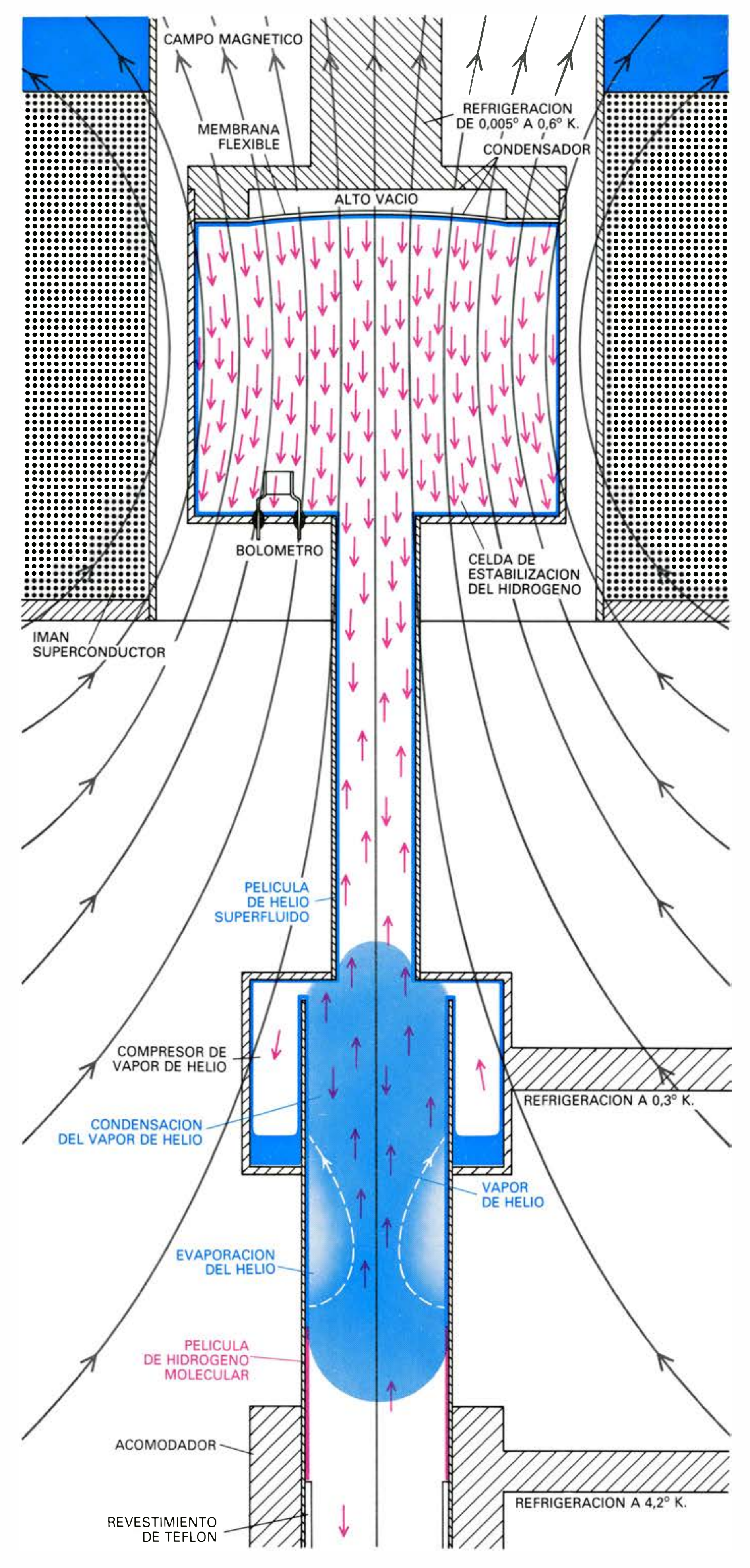
La transición de 20 grados a un grado Kelvin resulta crítica, pues el recubrimiento de teflón no sirve ya para evitar la recombinación. En el siguiente estadio de enfriamiento se necesitará otra clase de superficie, que disponga de un potencial de adsorción, o tendencia a capturar átomos de hidrógeno, más bajo. El helio, que cubre las paredes más frías de la celda de estabilización, tampoco vale en el estadio de transición. A esas temperaturas no formará una película superficial, a no ser que la presión de vapor del helio sea bastante elevada. Y no puede aceptarse una presión de vapor alta porque los átomos de helio catalizarían la rápida recombinación del gas de hidrógeno.

Se sabe que el mejor recubrimiento de una superficie, después del helio, es el hidrógeno molecular; se trata, además, de un sustituto aceptable para el helio en la región de transición. Stuart J. B. Crampton y sus colaboradores, del Williams College, demostraron que los átomos de hidrógeno no se adsorbían fuertemente en superficies sólidas de hidrógeno molecular a temperaturas por encima de los cuatro grados Kelvin. En nuestro aparato, se forma automáticamente una capa sólida de hidrógeno molecular cuando el hidrógeno atómico se recombina sobre la superficie de un tubo metálico llamado acomodador, que se mantiene a 4,2 grados Kelvin. Tras la introducción inicial de hidrógeno atómico, la acumulación decae a medida que la superficie de hidrógeno molecular inhibe nuevas recombinaciones.

La formacion es máxima en las regiones mantenidas entre 10 grados y un grado Kelvin y depende del tiempo durante el que se permite que el hidrógeno atómico permanezca en la región. Sabido eso, nuestra estrategia se aprestó a ajustar la longitud de esta región de modo que los átomos sufrieran un número de colisiones suficientes para enfriarse, pero no bastantes para recombinarse. El hidrógeno atómico que no se recombina en esta región alcanza finalmente la seguridad de la región cuyas paredes están recubiertas de helio.

Es en este último estadio del tratamiento del gas enfriado donde los átomos se separan por acción del gradiente de campo magnético según la dirección del spin del electrón. Los átomos con spin hacia abajo entran en la celda de estabilización del hidrógeno; y allí se mide la densidad final del hidrógeno atómico. La celda de estabilización se mantiene a temperaturas entre 0,5 y 0,1 grados Kelvin y se halla inmersa en un campo magnético de gradiente elevado que atrapa el gas atómico.

Antes de inyectar el hidrógeno, se introduce gas helio en la celda de estabilización enfriada, donde se condensa en la superficie en forma de helio 4 superfluido. La presión de vapor del helio re-



manente es despreciable a tan bajas temperaturas; y así la celda se evacuaría con la eficacia deseada. Una vez logrado un recubrimiento superficial de helio 4, se le saca partido a un efecto de la superfluidez. En efecto, el helio superfluido se expande por la superficie y hacia fuera de la celda hasta las regiones más calientes del acomodador. Allí se vaporiza; el vapor de helio tiende a volver a la celda en virtud de la presión más baja que registran las regiones más frías. Cuando el vapor se enfría, se condensa de nuevo en la superficie y se transforma en un superfluido.

La condensación del vapor de helio crea un difícil problema de refrigeración. Si se permite que el helio se condense en la celda de estabilización del hidrógeno, el calor de vaporización adquirido en la región más caliente se desprenderá en las paredes de la celda. La energía calorífica puede adquirir unos valores tan altos que impida alcanzar las bajas temperaturas que se exigen en la estabilización. ¿Cómo resolvimos el problema? Diseñando un estadio adicional de enfriamiento entre el acomodador y la celda de estabilización del hidrógeno. El aparato intermediario hace que el gas helio que vuelve se torne a condensar antes de alcanzar la celda y absorba así el calor de condensación. El gas helio que retorna ayuda también a confinar el gas hidrógeno en la celda. Los átomos de hidrógeno, cuyas trayectorias podrían permitirles escapar de la celda, chocan con los átomos de helio, de masa mayor, que se mueven en dirección contraria y son devueltos a la celda. De esta forma, el proceso de enfriamiento actúa como un compresor, impulsado por el vapor, para el gas hidrógeno polarizado.

NUCLEO del aparato experimental. Conduce hidrógeno atómico no polarizado y frío hacia una celda de estabilización en el seno de un campo magnético intenso. Cuando los átomos de hidrógeno se acercan al campo, la variación de la intensidad magnética repele los átomos con spin hacia arriba (*flechas apuntando hacia arriba*) y atrae los átomos con spin hacia abajo (*flechas apuntando hacia abajo*). Uno de los principales objetivos del diseño es evitar que los átomos de hidrógeno se peguen a las paredes del aparato, donde es más fácil que se recombinen. En la parte inferior y más caliente del aparato los átomos se enfrían hasta unos cuatro grados Kelvin al chocar con las paredes recubiertas de hidrógeno molecular (*rojo*). En la parte más fría y superior del aparato, las superficies están recubiertas de una película delgada de helio superfluido (*azul*) que tiene una tendencia mucho más baja a adsorber el hidrógeno atómico que la que tendría el hidrógeno molecular sólido. La densidad del gas de hidrógeno atómico puede determinarse bien a través de un condensador, que actúa como medidor de presiones, o mediante un bolómetro, instrumento que registra cambios en la temperatura como cambios en una resistencia eléctrica.

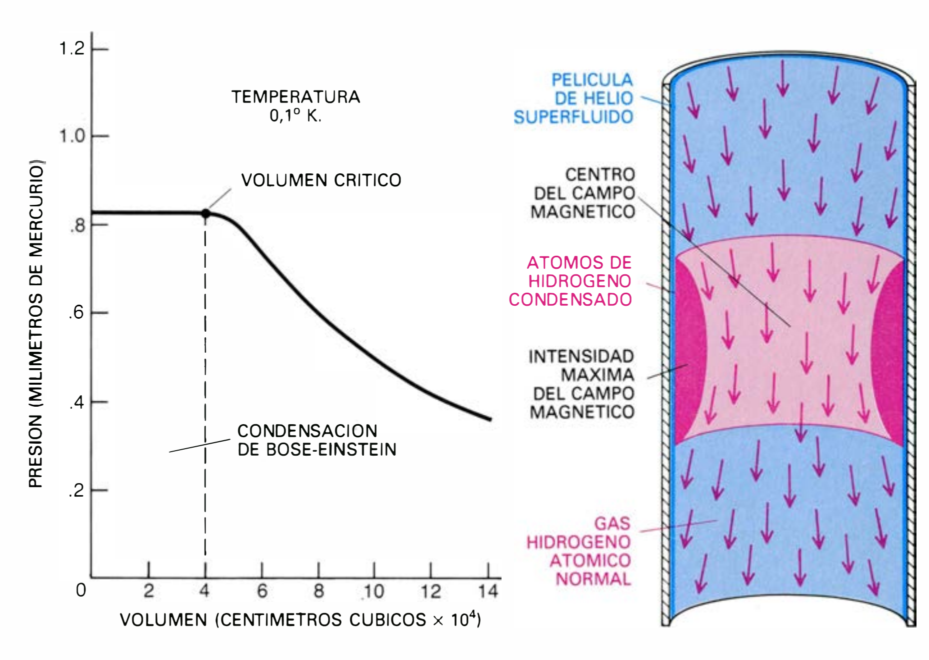
Desarrollamos luego una técnica sencilla y de sorprendente sensibilidad para detectar el hidrógeno atómico estabilizado. Cuando se recombinan dos átomos de hidrógeno, se desprende una gran cantidad de energía térmica. Haciendo que el gas que hay en la celda de estabilzación se recombine de repente, puede medirse el aumento de temperatura de la celda. Cabe, pues, determinar la densidad de hidrógeno atómico antes de la recombinación a partir de la energía de recombinación, que es conocida, y del volumen de la celda usada.

Para inducir la recombinación repentina y llevar a cabo la medición, instalamos una viruta cuadrada de carbón, de un milímetro escaso de lado, en la región de estabilización. Habíamos sacado la viruta de una resistencia de radio y la suspendimos en la celda sirviéndonos de cables finos. Puesto que la resistencia eléctrica del carbón cambia con la temperatura, la viruta funciona como un bolómetro; los cambios en la temperatura pueden calibrarse así a partir de los cambios en la resistencia. Mientras la celda de estabilización se llena de gas de hidrógeno atómico, los cables finos y la viruta de carbón se revisten de una película de helio superfluido.

Cuando quiere medirse la densidad del gas, se hace pasar una corriente eléctrica a través del bolómetro, calentándolo. El helio superfluido se evapora de la viruta a una tasa superior a la que pueda llenarla el flujo superfluido que discurra por los cables; así, la superficie desnuda del carbón queda expuesta al gas de hidrógeno atómico. No pasan 20 milisegundos sin que los átomos de hidrógeno de la celda choquen con la superficie del carbón y se recombinen, provocando una pequeña explosión en la celda y un aumento repentino y medible de su temperatura.

Pero el bolómetro, detector eficaz y barato, destruye la muestra de gas. Inconveniente que no presenta un medidor de presión de tipo condensador. La carga eléctrica que puede almacenarse en las dos placas paralelas de un condensador depende de la distancia entre ellas. Si una de las placas es una membrana sensible, las distorsiones de la membrana causadas por cambios en la presión podrán registrarse como señales eléctricas.

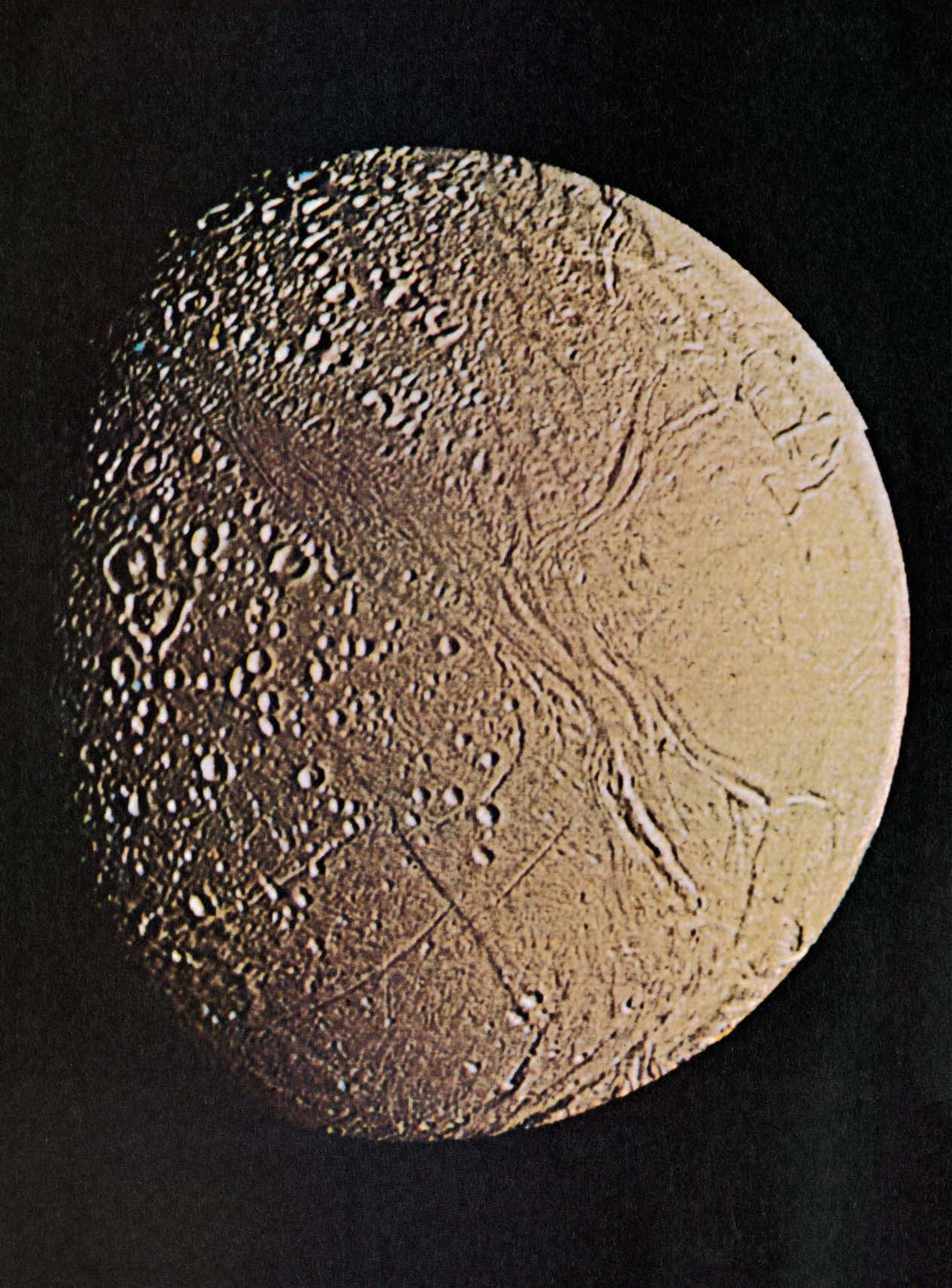
Consideradas las combinaciones de temperatura y densidad obtenidas hasta ahora con hidrógeno atómico ¿qué perspectivas hay de alcanzar la condensación de Bose-Einstein? Al intentar lograr temperaturas más bajas y densidades más altas, el primer problema que se planteó fue el siguiente: el revestimiento, por helio 4, de las paredes de la celda de estabilización del hidrógeno empieza a adsorber átomos de hidrógeno. La densidad alcanzable viene limitada, pues, por el aumento del ritmo de recombinación. Habría que hallar una superficie dotada de un potencial de adsorción más débil que helio 4. Se logra esto con una película delgada del isótopo helio 3 que flota sobre la capa de helio 4. Este revestimiento tiene un potencial de adsorción cifrado en torno a una tercera parte del de helio 4. Aunque una película de helio 3 permite aumentar la densidad del hidrógeno, no basta el incremento para permitir una condensación de Bose-Einstein.



CONDENSACION DE BOSE-EINSTEIN, que debería ser detectable estudiando la relación entre la presión y el volumen de un gas a temperatura constante. Para un gas ideal sin interacciones atómicas, la presión varía con el inverso del volumen, siempre que el volumen sea mucho mayor que el volumen crítico de condensación. Con la condición de volumen crítico, los átomos empiezan a ocupar el estado fundamental y dejan de contribuir a la presión del gas. No aumenta la presión si el gas se condensa más. Los átomos de hidrógeno que hay en el gas tienen momento magnético; significa ello que podemos localizarlos en el espacio con la ayuda del gradiente de un campo magnético. Los átomos en los estados de menor energía tienden a ocupar las regiones en las que el campo magnético posee máxima intensidad, de forma que los átomos del condensado Bose-Einstein deberían tender a amontonarse en las regiones de campo más fuerte. La región más intensa del campo generado por un solenoide se ubica en los lados del plano central del imán. Por consiguiente, si el hidrógeno gaseoso queda confinado en la parte interior del imán, los átomos condensados formarán, cabe esperarlo, un anillo denso en el plano central. (Dibujo de Gabor Kiss.)

Pero la investigación al respecto está tomando otra dirección. Un análisis detallado, fundado en la teoría cuántica, nos revela que el gas atómico de spin polarizado está formado por dos estados atómicos. En el estado *B* tanto el spin del protón como el del electrón están dirigidos hacia abajo. El estado *A* no se comporta de modo tan sencillo. En el estado *A* la configuración de los vectores de spin depende de la intensidad del campo magnético; para todas las intensidades finitas, el estado es una "mezcla" de las configuraciones. Aun cuando sea probable que el vector de spin del electrón para un átomo en el estado *A* esté dirigido hacia abajo, hay también una probabilidad pequeña, aunque no nula, de que esté hacia arriba; el vector spin del protón se halla en sentido opuesto al del electrón.

La mezcla de vectores de spin para los átomos en estado *A* concede alguna probabilidad de recombinación, incluso entre átomos de spin polarizado. Si los átomos del estado *B* pueden separarse de los del estado *A*, no cabe recombinación entre los átomos en el estado *B*. Richard W. Cline, Thomas J. Greytak y Daniel Kleppner, del Instituto de Tecnología de Massachusetts, acaban de preparar una muestra extremadamente pura de hidrógeno con spin polarizado en el estado *B*. Con un gas integrado de tales átomos podríase mantener mucho más elevada la concentración de hidrógeno atómico sobre una superficie antes de que el ritmo de recombinación se torne inaceptable. Cabría, entonces, comprimir el gas a las densidades exigidas para una condensación de Bose-Einstein. Si algún día se llegara a esto, la comunidad científica se encontraría con un estado de la materia que a buen seguro presentaría nuevas y exóticas propiedades cuánticas.



Los satélites de Saturno

Presentan una sorprendente variabilidad en su evolución geológica. Si bien muchos muestran cráteres de más de cuatro mil millones de años de antigüedad, hay uno cuya superficie, recentísima, carece de ellos

Laurence A. Soderblom y Torrence V. Johnson

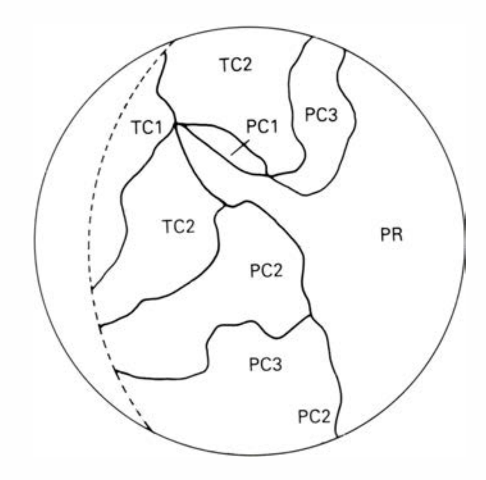
Antes de que el vehículo espacial *Voyager 1* se aproximase a Júpiter, en marzo de 1979, sólo se habían observado cinco cuerpos del sistema solar, aparte de la Tierra, con detalle suficiente para poder deducir su historia: Mercurio, la Luna y Marte y sus satélites Phobos y Deimos. Todos ellos están constituidos, esencialmente, de materia rocosa. El encuentro con Júpiter y sus satélites duplicó la lista. Además, señaló la primera aparición, en ese inventario, de satélites de tamaño planetario compuestos principalmente de hielos.

Vino luego la aproximación al sistema de Saturno, y, con ella, nueva duplicación del elenco. En noviembre de 1980, el *Voyager 1* pasó frente a Titán, el mayor satélite de Saturno, a una distancia de sólo 7000 kilómetros. Se acercó hasta los satélites menores Mimas, Dione y Rhea; aunque se mantuvo a una distancia mayor siguió transmitiendo a la Tierra imágenes de alta resolución de los mismos. En el proyecto del viaje constaba ya que la trayectoria del *Voyager 2* se aproximara, más de lo que lo había hecho el *Voyager 1*, hasta los satélites Iapetus, Hyperion y Phoebe; además, pasaría muy cerca de Tethys y Enceladus. En los meses que precedieron a la llegada del *Voyager 2* a la vecindad de Saturno, en agosto del año pasado, se alteró la secuencia de observaciones programada para el vehículo espacial a fin de reservar algunas observaciones para los satélites recientemente descubiertos, tres de los cuales lo habían sido por el *Voyager 1*. Pese a un agarrotamiento temporal en el mecanismo que sirve para apuntar las cámaras del *Voyager 2*, la misión tuvo éxito. En breves meses, pues, hemos asistido a la transformación de nuestro conocimiento de los satélites de Saturno. Antes de noviembre de 1980, sólo teníamos noticia de los que, a modo de puntos luminosos, se divisaban en los telescopios. Hoy forman un conjunto que da cabida a 17 mundos nuevos.

Propiedades generales

Se pueden sacar varias generalizaciones acerca de los satélites de Saturno. En primer lugar, sólo hay uno dotado de una atmósfera apreciable. Se trata de Titán, cuya atmósfera es opaca a la luz visible. Como nadie ha visto la superficie de Titán, la información directa que tenemos de su evolución geológica es mínima. En segundo lugar, se puede calcular que todos los satélites, salvo los tres más externos, deben ciertamente estar en rotación sincrónica: presentarán la misma cara vuelta hacia el planeta, de la misma manera que la Luna presenta siempre la misma cara vuelta hacia la Tierra. En cada uno de esos casos, la gravitación del planeta origina un abultamiento del satélite debido a la marea. La atracción gravitatoria entre el abultamiento y el planeta actúa entonces como un par que retarda la rotación del satélite hasta que ésta se hace sincrónica. Del examen de las imágenes del Voyager se desprende que todos los satélites de Saturno giran de forma sincrónica, menos uno o posiblemente dos. La excepción que no admite dudas es Phoebe, el satélite más externo, muy pequeño y demasiado alejado para perder su velocidad de rotación por causa de las fuerzas de marea. Gira una vez cada nueve horas, mientras que su período orbital es de 13.211 horas, es decir, 1,5 años. La excepción que hemos clasificado como posible es Hyperion, el tercer satélite contando desde fuera. Las imágenes de Hyperion hechas por el *Voyager 2* cubren sólo un pequeño arco de la órbita del satélite, y no aclaran si gira o no sincrónicamente.

Anotemos, por último, que todos los satélites de Saturno, salvo dos, constituyen un sistema regular. Significa ello que sus órbitas son casi circulares y quedan en el plano ecuatorial del planeta. Las dos excepciones son Iapetus, segundo satélite contando desde fuera, cuya órbita está inclinada 14,7 grados con respecto al plano ecuatorial, y Phoebe, el más externo, cuya órbita está inclinada 150 grados. (Además, el movimiento orbital de Phoebe en torno a Saturno tiene lugar en sentido opuesto al de todos los otros satélites.) En el sistema solar se conocen tres sistemas regulares de satélites: están constituidos por los satélites interiores de Júpiter, los satélites interiores de Saturno y los cinco satélites conocidos de Urano.

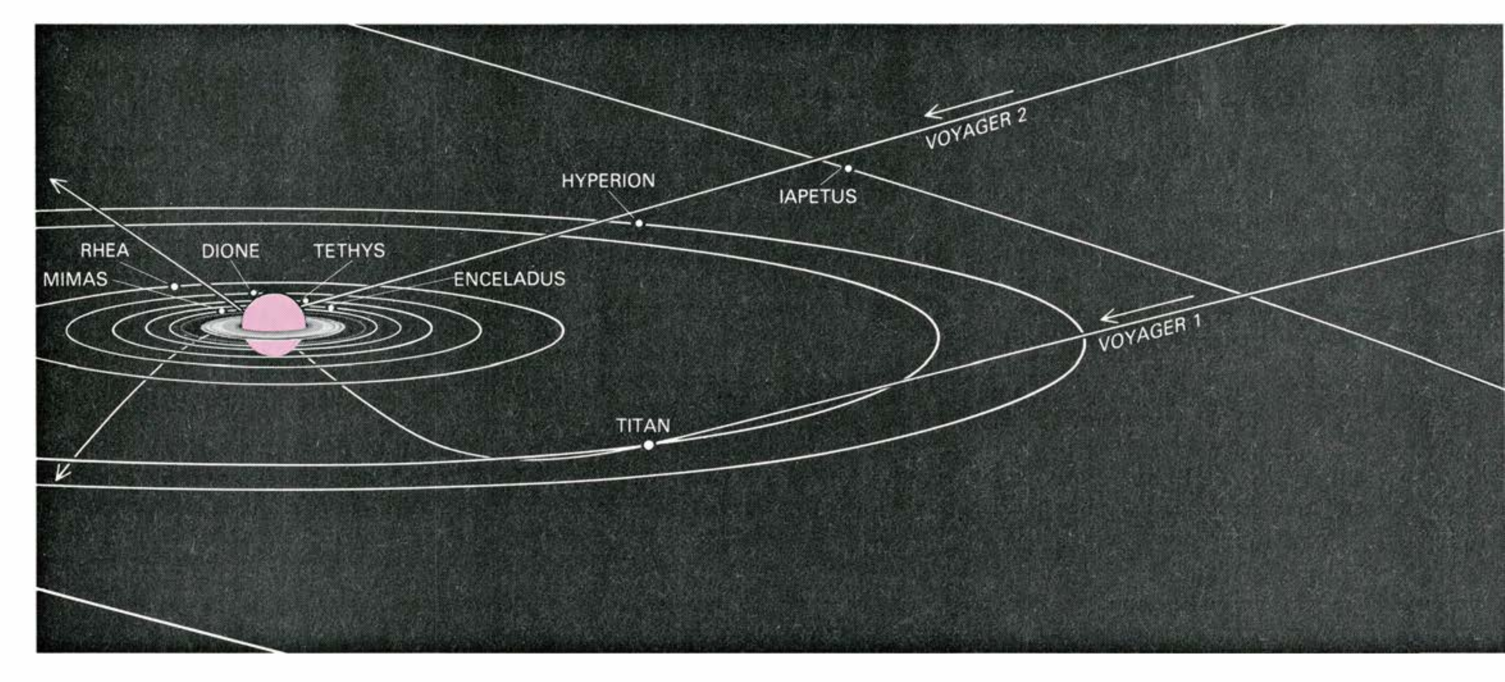


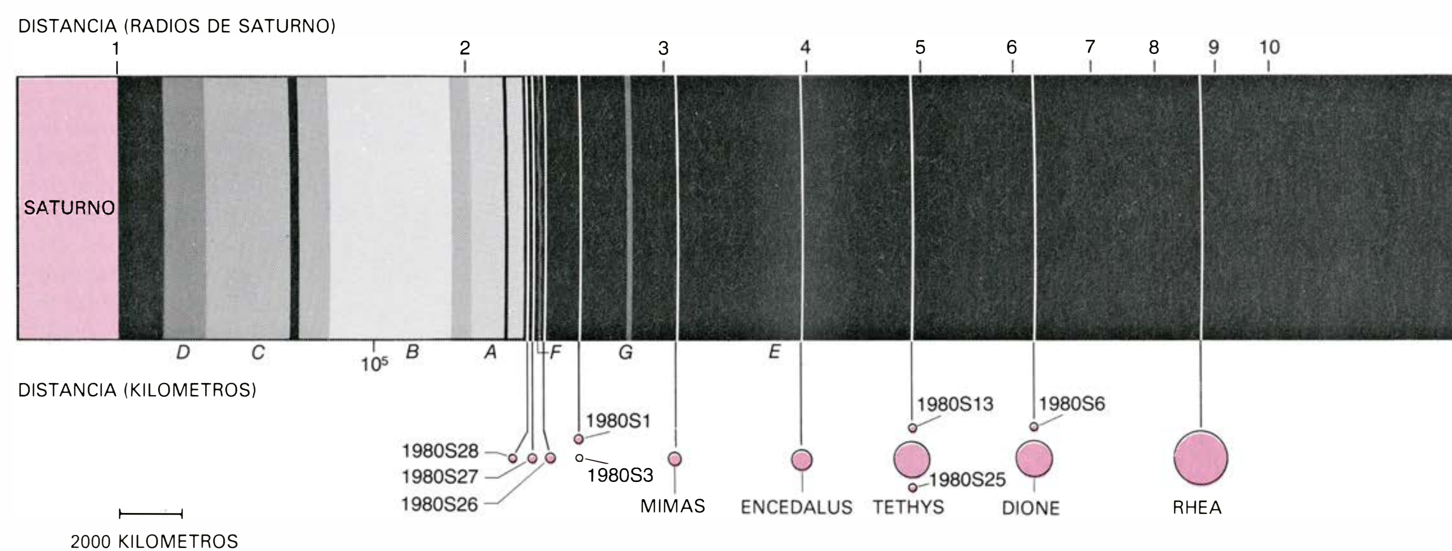
ENCELADUS, satélite de Saturno fotografiado el pasado 25 de agosto por el vehículo espacial *Voyager 2*. Mide unos 500 kilómetros de diámetro. Se ve aquí, a una distancia de 119.000 kilómetros, en una imagen que ha sido procesada por ordenador de suerte tal que se destaque la topografía de la superficie. Se distinguen seis clases diferentes de terrenos. Los más antiguos son los terrenos densamente cubiertos de cráteres llamados *TC*1 y *TC*2. Se cree que muchos de sus cráteres representan el bombardeo, sufrido por Enceladus y otros satélites saturnianos, de restos de materia en órbita alrededor de Saturno. Los restos persistieron tras la aglomeración de la materia para formar el planeta, sus satélites y sus anillos. Las planicies o llanuras salpicadas de cráteres *PC*1, *PC*2 y *PC*3 tienen una edad intermedia. La planicie *PR*, atravesada por sierras, es la más moderna; según una de las hipótesis, en ella no hay cráteres porque consta de materia que afloró en época bastante reciente.

Quizá no sea mera coincidencia el que todos esos planetas tengan anillos. Los anillos y un sistema regular de satélites pueden formarse naturalmente como un producto secundario de la deposición de materia sobre un planeta gigante. En cualquier caso, se cree que todo sistema regular de satélites se ha formado a partir del gas, hielo y polvo existente alrededor de cada planeta gigante incipiente, a semejanza de como, a mayor escala, nacieron los planetas alrededor del Sol.

La observación telescópica anterior a la llegada de los Voyager había mostrado que, entre los satélites de Saturno, sólo Titán poseía un tamaño comparable con los cuatro satélites de Júpiter descubiertos por Galileo; los otros eran menores que la Luna. Las imágenes de los Voyager revelan que los satélites abarcan un intervalo de tamaños desde el de los asteroides hasta el de Mercurio. Resulta prematuro especificar las masas de los satélites de Saturno. El análisis de los efectos gravitatorios mutuos que los satélites ejercen entre sí ha proporcionado valores para algunas de las masas; el seguimiento de las perturbaciones gravitatorias registradas en la trayectoria de los Voyager, a su paso por entre los satélites, aportó nuevos valores y refinó algunos de los anteriores. Partiendo de las distintas mediciones realizadas, parece claro que todos los satélites de Saturno tienen densidades inferiores a dos gramos por centímetro cúbico. Los hay incluso cuya densidad no llega al gramo y medio por centímetro cúbico. Tales valores sugieren una constitución principalmente de hielo. Para la mayoría de los satélites de Saturno, una composición de un 30 a un 40 por ciento de roca y un 60 a 70 por ciento de hielo en peso daría un resultado concordante con la densidad calculada. Sólo Titán posee magnitud suficiente para que su autocompresión gravitatoria afecte, de un modo apreciable, a su densidad. Cuando se tiene en cuenta esta autocompresión, la densidad estimada para Titán –1,9 gramos por centímetro cúbico– resulta compatible con una mezcla, a partes iguales, de roca y hielo.

Esas densidades calculadas tienen un punto de curiosidad. Sabido es que, en-



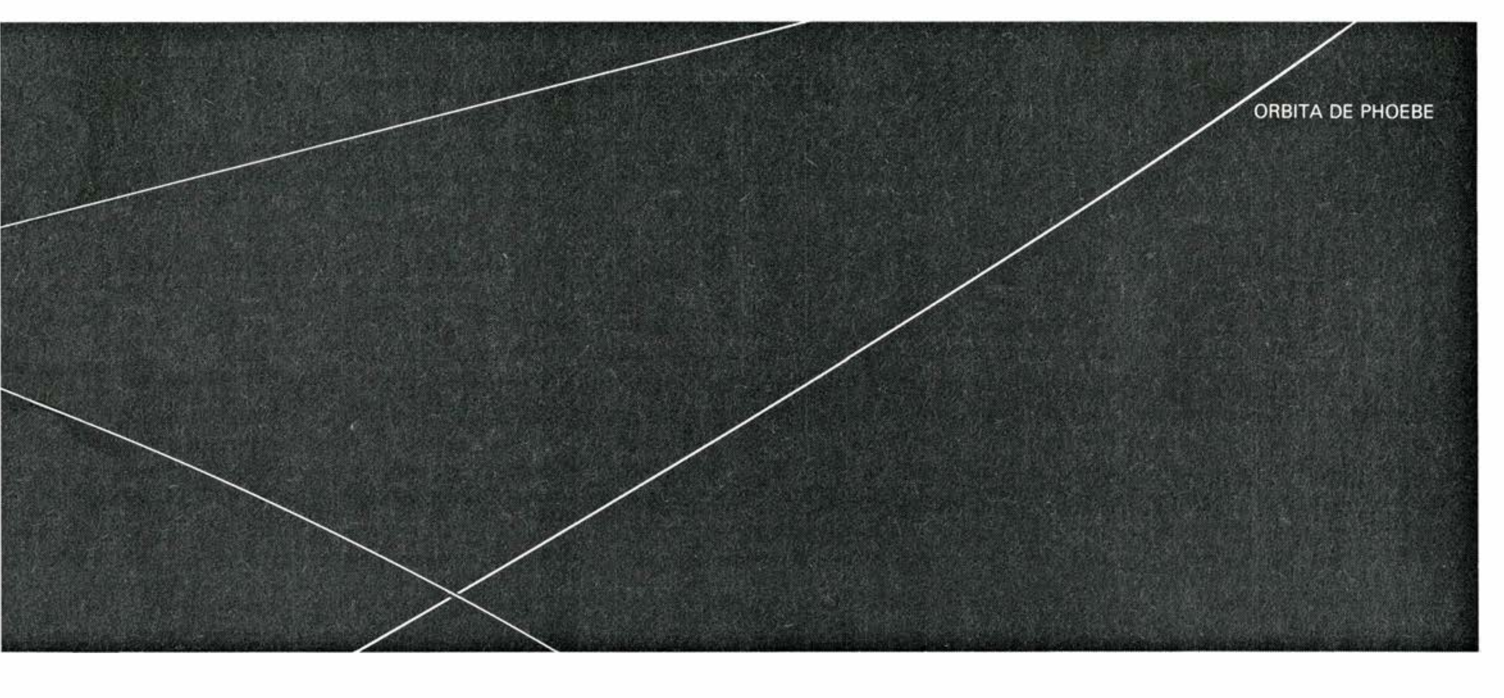


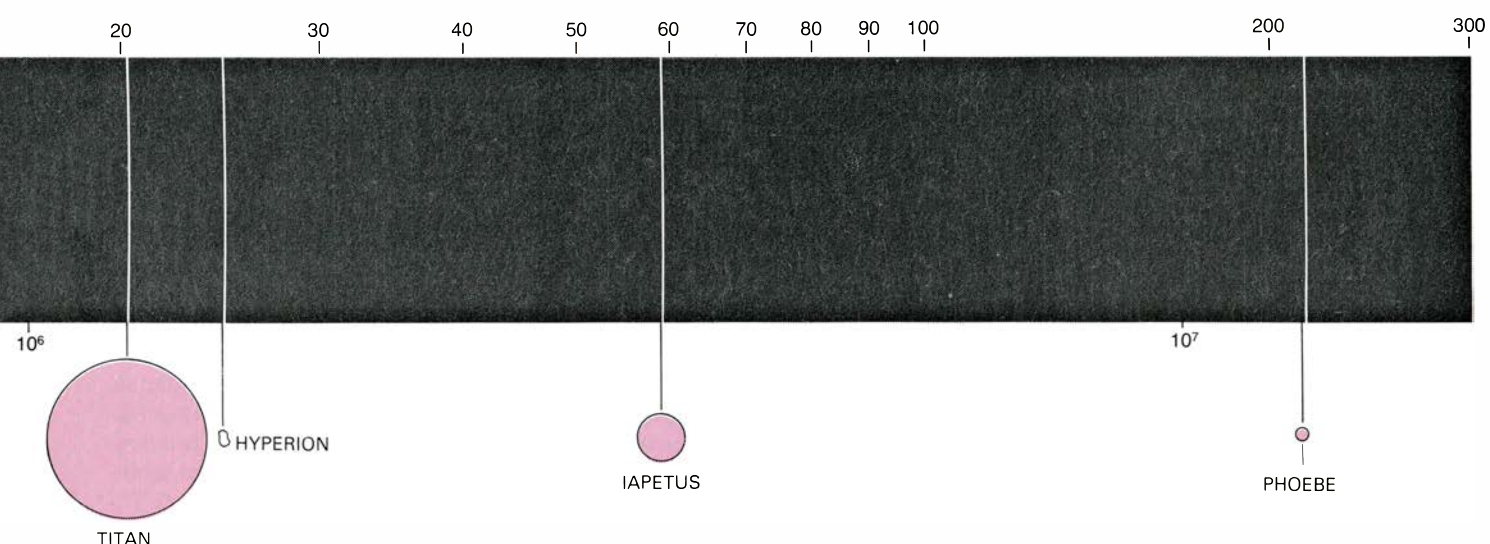
EXISTEN 17 SATELITES DE SATURNO conocidos. El dibujo superior recoge sus órbitas. A excepción de dos, el resto descansa en el plano ecuatorial del planeta, que es también el de los anillos del mismo. Iapetus (cuya órbita está inclinada 14,7 grados) y Phoebe (que la tiene inclinada 150 grados) constituyen las salvedades. La rotación de Phoebe alrededor de Saturno procede en sentido horario; los demás satélites giran en sentido antihorario. El dibujo superior muestra también las trayectorias seguidas por el *Voyager 1* en noviembre de 1980 y por el *Voyager 2* en el pasado agosto. El dibujo de la parte central ofrece las órbitas de los satélites en escala logarítmica; las cifras en la escala son las distancias desde el centro de Saturno. Todos los satélites descu-

tre los planetas del sistema solar, existe la tendencia hacia mayor densidad con distancia decreciente al Sol; fenómeno que ocurre también entre los satélites de Júpiter, donde se aprecia una tendencia hacia mayor densidad conforme se acorta la distancia al planeta. Tales tendencias se atribuyen a la influencia del calentamiento del cuerpo central sobre la temperatura del gas y polvo que lo rodeaban en la época en que se formaron los satélites. En el caso de Júpiter, parece que el hielo era inestable, en proporción muy elevada, a distancias radiales del planeta inferiores a la distancia de la órbita que ahora sigue Ganimedes.

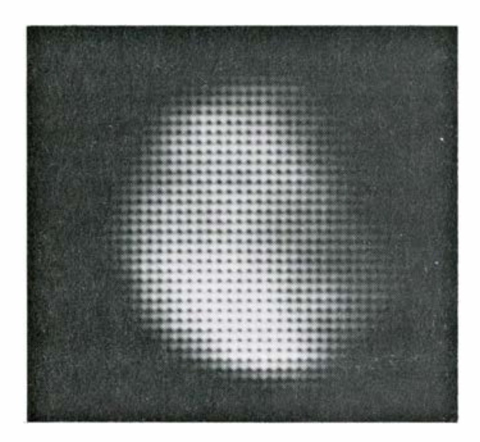
A pesar de la falta de precisión en las mediciones tomadas, no advertimos una tendencia similar de las densidades entre los satélites de Saturno. Por el contrario, las determinaciones actuales de los valores de las densidades concuerdan con una composición de roca y hielo que es semejante en todos esos cuerpos, si pasamos por alto variaciones más o menos arbitrarias en las proporciones exactas. Por otro lado, las densidades de los satélites de Saturno son, en general, inferiores a las de los de Júpiter. Esto sugiere una mayor proporción de hielo. La relativa falta de materia rocosa halla una explicación en los modelos de la historia de Saturno desarrollados por James B. Pollack y sus colegas, del Centro de Investigación Ames de la Administración Nacional de Aeronáutica y del Espacio (NASA). Los modelos sugieren que el material rocoso en la vecindad de Saturno fue arrastrado hacia el planeta incipiente cuando éste se contrajo, antes de que sus satélites comenzaran a formarse hace unos 4500 millones de años.

En cualquier caso, la superficie de los satélites de Saturno sugiere la presencia de hielo. Los espectros de la radiación solar reflejada por ellos en el infrarrojo próximo muestran su absorción a longitudes de onda características del agua en estado sólido. Además, las medidas hechas por los Voyager muestran que la mayoría de los satélites reflejan entre el 60 y 90 por ciento de la radiación incidente sobre ellos. Encelados es el cuerpo más reflector del sistema solar. Si estuviera a la misma dis-





biertos se distinguían por algún capricho dinámico. El satélite 1980S28 está situado cerca del borde exterior del anillo *A*; el 1980S27 y el 1980S26 comprenden entre sí el anillo *F*, y el 1980S1 y 1980S3 tienen órbitas que difieren en menos de la suma de los diámetros de los satélites (de aquí que sean "coorbitales"). Finalmente, 1980S13 adelanta en 60 grados a Tethys, 1980S25 está 60 grados retrasado respecto a Tethys y 1980S6 adelanta en 60 grados a Dione. Cada uno de los tres ocupa una posición estable, estudiada por Joseph Louis Lagrange en el siglo XVIII y que, en su honor, recibe hoy el nombre de punto de Lagrange. Se trata de los primeros satélites lagrangianos conocidos. En la parte baja aparecen detallados, a escala, todos los satélites de Saturno.



PHOEBE, el satélite más externo de Saturno, fue fotografiado por el *Voyager 2* a una distancia de 2.200.000 kilómetros. Su curiosa órbita y su baja reflectividad (alrededor del 5 por ciento) sugieren que se formó en otra parte del sistema solar y cayó después capturado por la atracción gravitatoria de Saturno. Tiene unos 200 kilómetros de diámetro.



HYPERION, el tercer satélite contando desde fuera, fotografiado por el *Voyager 2* a una distancia de 500.000 kilómetros. Se trata de un cuerpo irregular, de unos 400 kilómetros de largo por 220 de ancho. Brilla cuatro o seis veces más que Phoebe. Las concavidades a lo largo del borde de esta imagen de Hyperion son probablemente cráteres.

tancia del Sol que la Luna, brillaría unas cinco veces más. El hielo el altamente reflector.

La ubicuidad del hielo en el sistema solar exterior no nos tomó por sorpresa. Por una parte, la presión de vapor del hielo (es decir, la tendencia del hielo a sublimarse y perder vapor hacia el espacio) depende estrechamente de la temperatura. De aquí que, a distancias del Sol inferiores a la distancia del cinturón de asteroides, sito entre las órbitas de Marte y Júpiter, una masa de hielo no protegida se evaporará en un tiempo muy corto comparado con la edad del sistema solar. A distancias mayores, una masa de hielo mantendrá su estabilidad a lo largo de miles de millones de años. Téngase en cuenta, además, que la mayoría de modelos propuestos para explicar la formación del sistema solar predicen que el agua debe ser un componente importante de los cuerpos que se condensaron a bajas temperaturas.

Concretando: si un gas cuya composición sea muy semejante a la del Sol se enfriase bajo las condiciones de temperatura y presión que se cree existían en el primitivo sistema solar, parte del oxígeno del gas se combinará con el silicio para formar rocas silíceas a temperaturas bastante altas. Cuando el silicio elemental se haya agotado, quedará todavía una proporción importante de oxígeno. Al continuar decreciendo la temperatura, se combinará con el hidrógeno, el elemento más abundante en el gas. Se formará, pues, agua. De tales modelos resulta que un satélite que se condensase a temperaturas bajas debería tener proporciones de roca y hielo muy semejantes a las deducidas, para los satélites de Júpiter y Saturno, a partir de las estimas de sus densidades.

Alguien podría afirmar que el interés geológico de un satélite de hielo no trasciende el de un simple cubito de hielo salpicado de cráteres. Pero el hielo tiene un punto de fusión mucho más bajo que el de la roca; se necesita relativamente poco calor para fundir el interior de un satélite del sistema solar exterior que contenga hielo. Sobre esta base, se sospechaba, ya antes de las misiones Voyager, que tales satélites podrían mostrar señales de actividad geológica. Se proponía también que los satélites del sistema solar exterior quizás incorporaran ciertas sustancias: hidratos amónicos y los compuestos de metano y agua conocidos como clatratos. Mas el interior de un satélite que incluya tales materias acabaría por fundirse mucho antes que un satélite formado sólo de roca y hielo. Los resultados de las misiones Voyager superarían las especulaciones más arriesgadas. Hyperion, Mimas y Enceladus, por ejemplo, tienen un tamaño muy parecido, pero presentan una variación de evolución geológica mucho más amplia de lo que se podía considerar probable.

Los nuevos satélites

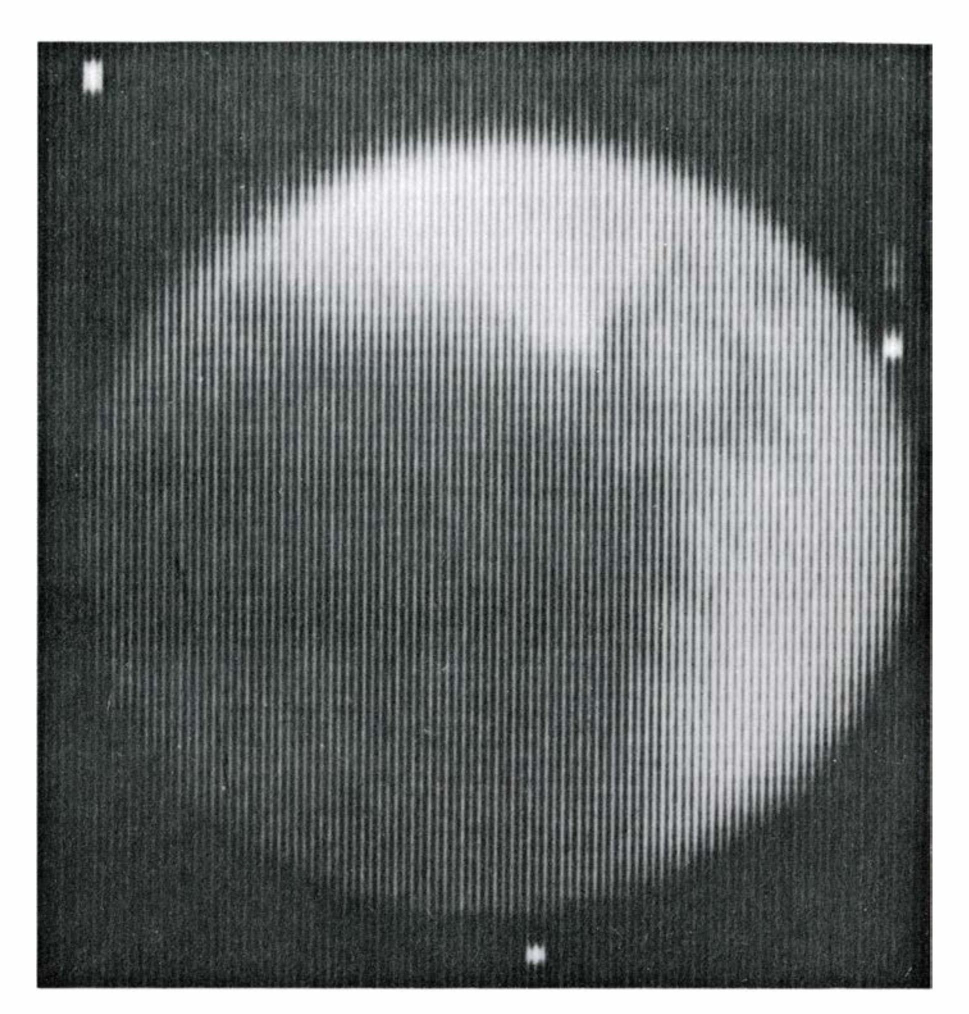
Los satélites de Saturno descubiertos en el pasado decenio presentan dinámicas que, de una u otra manera, resultan poco corrientes. Considérese 1980S28, el satélite más interno del conjunto y el más interno también de los conocidos en Saturno. Fue descubierto por el *Voyager 1* inmediatamente más allá del borde exterior del anillo *A* del planeta. 1980S28 es un cuerpo alargado cuyo diámetro mide unos 40 kilómetros. Una hipótesis sugiere que su propio campo gravitatorio contribuye en gran parte a esculpir el abrupto borde exterior del anillo. Algo más hacia fuera de Saturno se encuentra la pareja de pequeños satélites 1980S27 y 1980S26. Debemos su descubrimiento también al *Voyager 1*, aunque ahora reconozcamos que uno al menos del par pudo afectar al recuento de partículas cargadas realizado por el *Pioneer 11* cerca de Saturno en septiembre de 1979. Entre sus órbitas queda el anillo *F*, de varias bandas. Los campos gravitatorios de los satélites pueden muy bien confinar el anillo. Por cuya razón se les denomina satélites pastores.

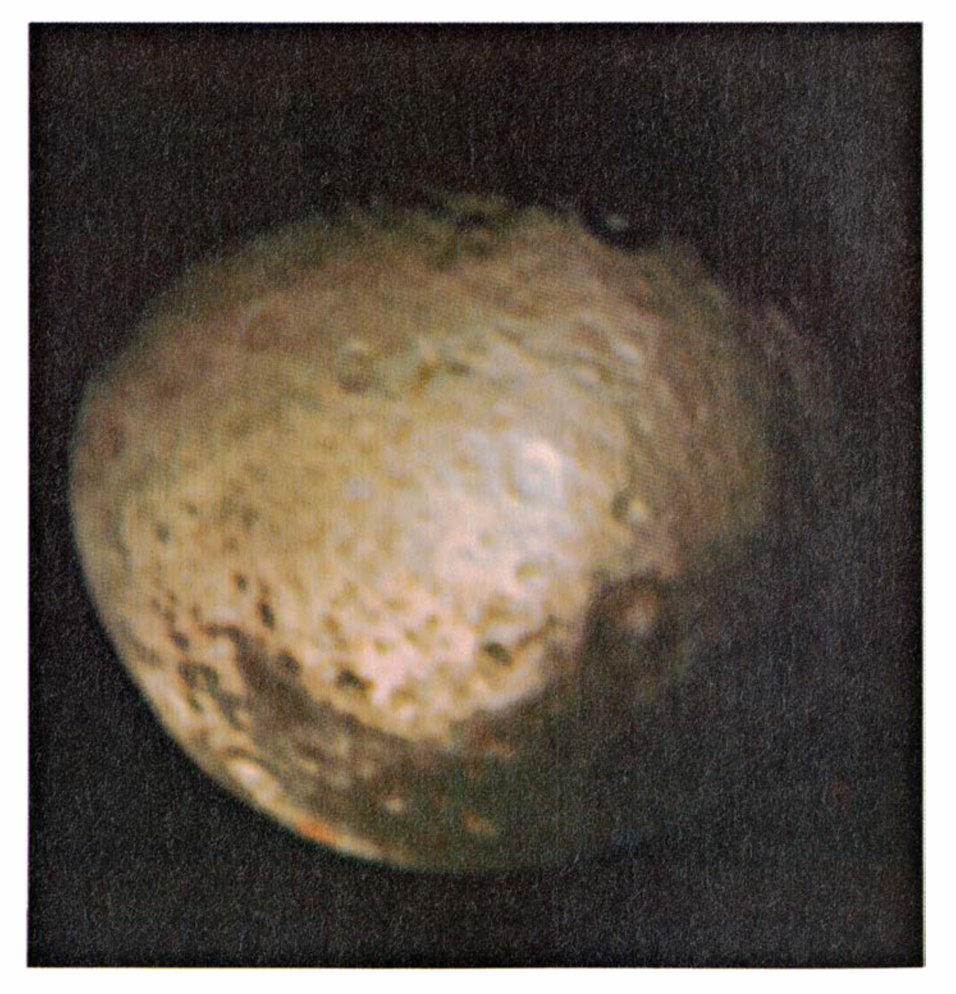
A unos 10.000 kilómetros hacia el exterior del anillo *F*, lo que viene a ubicarse a medio camino entre los anillos *F* y *G*, están los satélites "co-orbitales" 1980S1 y 1980S3. El astrónomo francés Audouin Dollfus fotografió uno de ellos en el Observatorio del Pic du Midi en 1966. Es difícil decidir hoy cuál de ellos era. En 1978, John W. Fountain y Stephen M. Larson, de la Universidad de Arizona, determinaron que se trataba de dos satélites. Después, en 1979, el *Pioneer 11* tomó una imagen de uno de ellos. Los Voyager las tomaron de ambos. Los radios orbitales medios de 1980S1 y 1980S3 difieren en menos de la suma de sus diámetros. De ahí que sus velocidades orbitales se asemejen, pero no sean del todo idénticas: el satélite interior adelanta lentamente al exterior. Cuando se acercan entre sí, la mutua atracción gravitatoria altera sus momentos de cantidad de movimiento. El satélite interior gana cantidad de movimiento; se mueve hasta una órbita mayor, en la que su velocidad orbital se reduce. El satélite exterior pierde cantidad de movimiento; se traslada hasta una órbita menor, donde su velocidad orbital aumenta. En resumen, los dos satélites intercambian sus lugares; el satélite interior se convierte en exterior y comienza a rezagarse. La danza celestial se repite aproximadamente una vez cada cuatro años: los satélites tornan a intercambiar entonces su puesto.

Entre 1979 y 1980, época en la que los anillos de Saturno se veían de canto desde la Tierra, los telescopios afincados en ésta hallaron otros tres nuevos satélites. Esa orientación de los anillos reduce grandemente la luz que difunden cuando contemplamos el planeta desde un telescopio y permite la detección de débiles cuerpos en la proximidad de éste. De los tres nuevos satélites, dos comparten su órbita con Tethys: uno, el descubierto por un grupo dirigido por Bradford A. Smith, de la Universidad de Arizona, mantiene una posición de unos 60 grados delante de Tethys; el otro queda a unos 60 grados detrás de Tethys; fue descubierto con la

ayuda de un prototipo de un sistema de cámara planetaria proyectada para el telescopio que los Estados Unidos piensan poner en órbita terrestre. El tercer satélite queda a unos 60 grados delante de Dione; lo descubrieron dos astrónomos franceses, P. Lacques, del Observatorio del Pic du Midi, y J. Lecacheux, del Observatorio de París.

Las posiciones ocupadas por esos satélites constituyen puntos de estabilidad dinámica de la clase estudiada, en primer lugar por el matemático francés Joseph Louis Lagrange; lo que se llaman hoy puntos de Lagrange. Observó éste, en 1772, que en un sistema consistente en un cuerpo en órbita alrededor de otro (por ejemplo, la Luna alrededor de la Tierra), había cinco posiciones en las que un tercer cuerpo podía quedar imperturbado. Tres de esas posiciones eran inestables: un cuerpo que estuviera en una de ellas resultaba fácilmente alejado de la misma por la influencia de fuerzas gravitatorias distintas de las ejercidas por los dos cuerpos que constituían el sistema. Las tres posiciones estaban situadas: (1) dentro de la órbita del satélite, (2) fuera de la órbita y (3) en el punto de la órbita opuesto al satélite mismo. Las dos posiciones restantes –las que estaban en la órbita 60 grados delante del satélite y 60 grados detrás– eran muy estables: un cuerpo que ocupara cualquiera de ellas derivaría simplemente hacia atrás y hacia adelante en su órbita bajo la influencia de fuerzas perturbadoras. Se sabe, desde hace mucho tiempo, que los grupos de asteroides llamados Troyanos ocupan los dos puntos de Lagrange a 60 grados de Júpiter en la órbita del planeta alrededor del Sol. Los tres cuerpos en puntos de Lagrange





IAPETUS fue fotografiado por el *Voyager 1* a una distancia de 3.200.000 kilómetros, y nueve meses más tarde por el *Voyager 2* a una distancia de 1.100.000 kilómetros. La imagen sacada por el *Voyager 1* (*arriba*) revela la cara del satélite que siempre mira a Saturno. (Iapetus y la mayoría de los restantes satélites de Saturno presentan siempre la misma cara mirando a Saturno, al igual que la Luna presenta siempre la misma cara vuelta hacia la Tierra.) Iapetus se mueve hacia la izquierda; su hemisfero anterior (el que mira en la dirección del movimiento orbital del satélite) es oscuro. Sin embargo, hay un anillo de materia oscura que se extiende hasta el hemisferio posterior. La imagen del *Voyager 2* (*abajo*) nos aporta la cara del satélite que siempre mira hacia la parte de fuera. El polo Norte coincide aproximadamente con el cráter grande, a caballo sobre el borde entre el día y la noche en Iapetus (el terminador), que cruza la parte superior de la imagen. El oscuro hemisferio anterior está a la derecha. Una banda ecuatorial oscura se extiende hacia el hemisferio posterior. La marcada y complicada frontera que divide la región oscura de la brillante contradice la hipótesis de que cayera materia oscura desde el espacio sobre el satélite. Pero apoya la idea de que dicha materia oscura saliera del interior de Iapetus.

cercanos a Saturno son los primeros satélites lagrangianos conocidos.

En contraste con sus peculiaridades dinámicas, los satélites recién descubiertos presentan una morfología afín. Bastante pequeños, tienen casi todos una forma irregular. La irregularidad enseña mucho. Sugiere, en efecto, que cada satélite nació a partir de la fragmentación de un cuerpo mayor. Y como un satélite pequeño, frío y formado de hielo es lo bastante fuerte para resistirse a adquirir la forma esférica en razón de su autograviatación, el dibujo irregular pone de manifiesto que esos satélites no han sufrido un calentamiento especial desde el momento en que adquirieron su actual configuración.

Iapetus

Iapetus y Rhea ocupan el segundo puesto en la jerarquía de tamaños en que podrían ordenarse los satélites de Saturno. Tienen un diámetro casi idéntico de unos 1500 kilómetros. Pero se mueven en órbitas muy diferentes. La de Iapetus está situada a unos 60 radios saturnianos (R_s) del centro de Saturno, y la de Rhea a unos 9 R_s. Entre sí encierran las órbitas de otros dos satélites de Saturno, Titán e Hyperion. Como hemos observado, la órbita de Iapetus está inclinada.

Iapetus tiene una densidad aproximada de 1,1. Lo que significa que es casi tan baja como la del agua pura. Por presentar la misma cara vuelta hacia Saturno, el hemisferio posterior (el hemisferio que mira atrás con respecto a su movimiento orbital) es siempre la misma parte de la superficie. Este hemisferio es brillante. En la región visible del espectro electromagnético, su albedo es casi del 50 por ciento. Lo que contrasta con lo que ocurre en el hemisferio anterior, extremadamente oscuro: su albedo es sólo del 3 al 5 por ciento. Esta diferencia fue ya observada, en el siglo XVII, por Jean Dominique Cassini, descubridor de Iapetus, quien halló que podía ver el cuerpo a un lado de Saturno, pero no al otro. Para hacernos una idea más cabal, albedo tan pequeñó sólo lo ofrecen el hollín y los meteoritos primitivos llamados condritas carbonáceas, y pocos más.

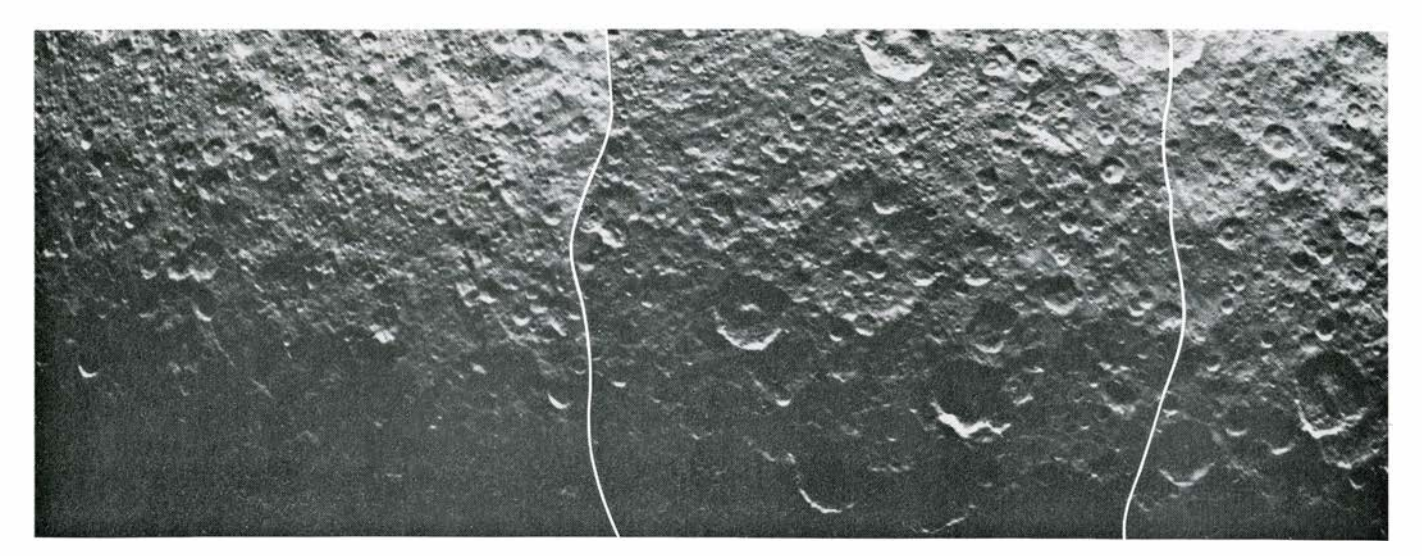
La peculiar configuración de oscuridad y brillo en Iapetus ha sugerido que algún tipo de materia oscura que caiga del espacio recubriría la porción que actúa a modo de proa del satélite. En 1974, Steven Soter, de la Universidad Cornell, llegó a especificar que la fuente de dicha materia sería Phoebe, satélite considerado, durante mucho tiempo, como escasamente reflector. Las observaciones, hechas con telescopios emplazados en la Tierra, muestran que su color semeja al de muchos asteroides oscuros. El razonamiento es el siguiente: la materia oscura salpica desde Phoebe en virtud del impacto de micrometeoritos; las partículas de esa materia caen, luego, bajo la influencia del efecto Poynting-Robertson. La radiación electromagnética abandona cada partícula porque ésta refleja parte de la radiación incidente sobre ella y absorbe otra, que reemite después. En cualquier caso, la radiación que procede de la partícula en la dirección del movimiento orbital de ésta en torno a Saturno sufre un corrimiento Doppler que le proporciona una frecuencia más alta (y consiguientemente, más energía y cantidad de movimiento) que la radiación que abandona la partícula en sentido opuesto. El resultado neto es que la partícula pierde momento de la cantidad de movimiento orbital y, lentamente, describe una espiral hacia dentro. Al caer, concluye la argumentación, es barrida por Iapetus.



RHEA fue fotografiada por el *Voyager 1* a una distancia de 1.700.000 kilómetros. Aquí se muestra la imagen del satélite en falso color. El hemisferio posterior que vemos está marcado por franjas brillantes que no parecen seguir la configuración impuesta por los cráteres. El hemisferio anterior es liso.

Las fotografías de Phoebe sacadas por el *Voyager 2* revelan ciertos detalles de su superficie. Aparece el satélite toscamente esférico, con un albedo de sólo un 5 por ciento. En ambos atributos (y en su órbita inclinada y retrógrada) difiere por completo de sus compañeros orbitales, los pastores de anillos y los satélites lagrangianos de Saturno. Cabe la posibilidad de hallarnos ante un asteroide oscuro capturado que no ha sufrido cambios desde que se formó en la época primitiva de la historia del sistema solar. Quizá sea un cuerpo primario que fue expulsado del sistema solar interior por el campo gravitatorio del creciente planeta Júpiter y quedó en órbita alrededor de Saturno.

Por sí solas, las imágenes de Phoebe que nos facilitaron los Voyager no pueden dejar en claro si Phoebe cubrió de polvo el oscuro hemisferio anterior de Iapetus. Se trata de una propuesta sometida todavía a discusión. En su comprobación algunos involucran a Hyperion, el satélite que sigue a Iapetus en el acercamiento a Saturno: se basa en que Hyperion, razonan, debería haberse cubierto de polvo por el material expulsado de Phoebe que Iapetus dejó de barrer. Las imágenes tomadas por el *Voyager 2* muestran que Hyperion destaca por su reducido tamaño, más o menos el de Mimas, satélite esférico de unos 200 kilómetros de radio.



POLO NORTE DE RHEA, fotografiado por el *Voyager 1* a la distancia de 80.000 kilómetros. Se halla en la parte central del arco que el terminador describe a través de este mosaico. El terreno al Oeste del polo (que cae hacia la derecha en esta imagen) está salpicado de grandes cráteres; en el terreno hacia el Este no hay tales accidentes. Los cráteres grandes se produjeron en la época primitiva de la historia del satélite; los cráteres pequeños, tanto al Este como al Oeste, se causaron más tarde, cuando la actividad geológica había dotado de nueva superficie al terreno que se extiende al Este del polo.

Se caracteriza por su contorno alargado: su radio menor es aproximadamente tres quintas partes del radio mayor. Es, además, uno de los mayores cuerpos del sistema solar con forma irregular. (El asteroide Héctor es otro de los contados objetos que integran dicha clase.) El análisis de las imágenes del Voyager indica que Hyperion es algo más oscuro que la mayoría de los satélites de Saturno, con un albedo del 20 al 30 por ciento. Este descubrimiento concuerda con la idea del empolvamiento por materia procedente de Phoebe. No parece, sin embargo, que haya de tener un hemisferio anterior oscuro. El problema de su búsqueda estriba en que todavía no está establecido si el satélite posee o no rotación sincrónica.

Así pues, la respuesta a la pregunta de si Phoebe empolvó o no Iapetus deberá aguardar, por lo menos, a que las observaciones realizadas con telescopios emplazados en la Tierra determinen la velocidad de rotación de Hyperion. No acabarán ahí las dificultades. Los descubrimientos de Dale P. Cruikshank y sus colaboradores, de la Universidad de Hawai en Manoa, pueden haber puesto la idea en un brete por otras razones. El grupo de Cruikshank ha hallado, a partir de observaciones telescópicas, que el color de Phoebe difiere del color del hemisferio anterior de Iapetus. Las regiones oscuras de éste son mucho más rojas que Phoebe en toda la región visible e infrarroja próxima del espectro. Se hace, pues, difícil aceptar un esquema en el cual la materia procedente de Phoebe simplemente se deposita recubriendo la parte oscura de Iapetus, sin haber sufrido alguna clase de cambio.

Las imágenes de Iapetus hechas por los vehículos espaciales Voyager ponen también en cuestión la hipótesis de que fue empolvado por materia procedente de Phoebe. La mejor imagen de Iapetus realizada por el *Voyager 1* se tomó a una distancia de 2,5 millones de kilómetros. Sólo mostraba aquellos accidentes del satélite que tenían diámetro superior a 50 kilómetros. Lo que no fue óbice para que salieran a luz algunas claves que nos acercaran a la historia del cuerpo. Se encontró que un anillo de materia oscura, de unos 100 kilómetros de diámetro, ceñía la frontera entre los hemisferios. Presentaba un gran parecido con los anillos de los cráteres situados en el margen de grandes llanuras de inundación volcánica de la Luna y Marte. Tales anillos se formaron allí cuando la oscura lava volcánica fluyó hacia el interior de los cráteres de impacto y los rellenó en torno al pico central. Quizá el anillo oscuro de Iapetus se formó por un proceso análogo, cuando el fluido rebosado del interior del cuerpo ennegreció la mitad de éste. Se comprende, pues, la improbabilidad de que un accidente con geometría tan peculiar como la de un anillo oscuro pudiera haberse formado por materia que cayese del espacio.

El *Voyager 2* tomó una serie de imágenes de Iapetus a una resolución tres veces mayor que la obtenida en las fotografías del *Voyager 1*. Sobresale, por su calidad, la relativa a la región polar septentrional, que, en su mayor parte, se extiende por el brillante hemisferio posterior. Nos revela cómo este hemisferio está intensamente salpicado de cráteres. Muchos de ellos tienen un fondo oscuro, parecido al también oscuro de los cráteres de las altiplanicies de la Luna terrestre, que se cree se formaron cuando la materia volcánica fluyó hasta su interior. Tomados juntos, los cráteres de fondo oscuro del hemisferio posterior y la marcada definición y complejidad entre el terreno oscuro y el brillante, implican una historia de erupciones desde el interior de Iapetus. Esto no quiere decir que el material de erupción se parezca a la lava ordinaria. Se trataba, presumiblemente, de una mezcolanza fluida que incluía amoniaco, hielo y algún material oscuro. El bajo albedo del material sugiere que, cualquiera que fuera su origen, la materia oscura es rica en sustancias órganicas que se encuentran en los primitivos meteoritos.

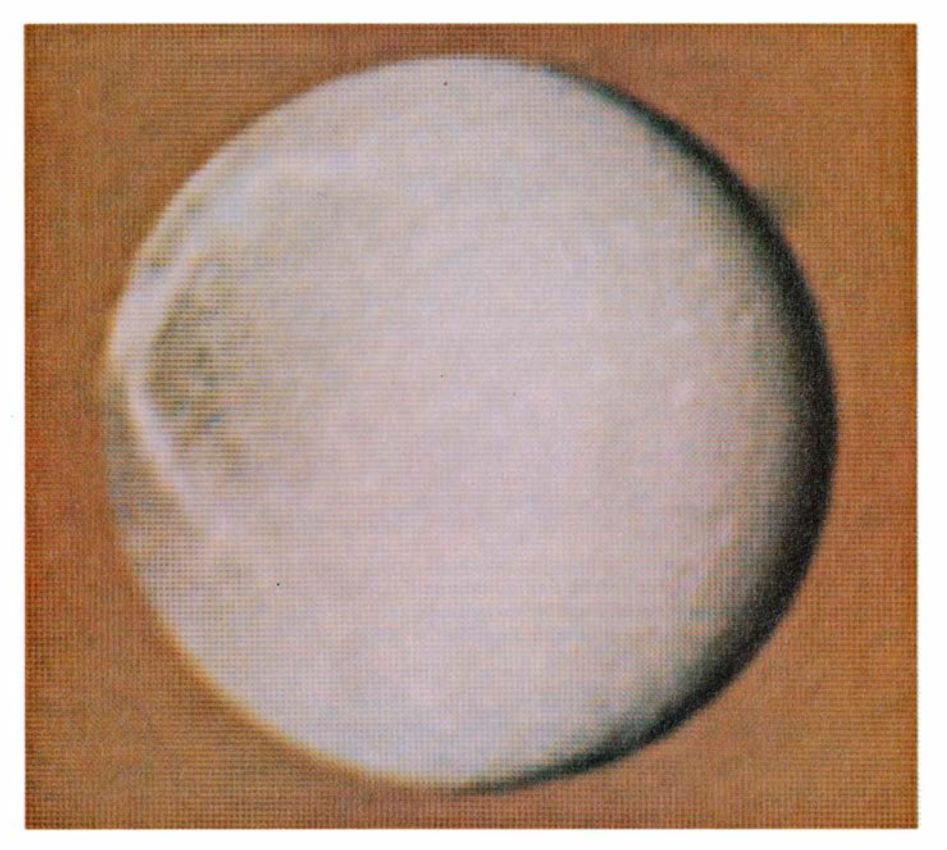
Rhea

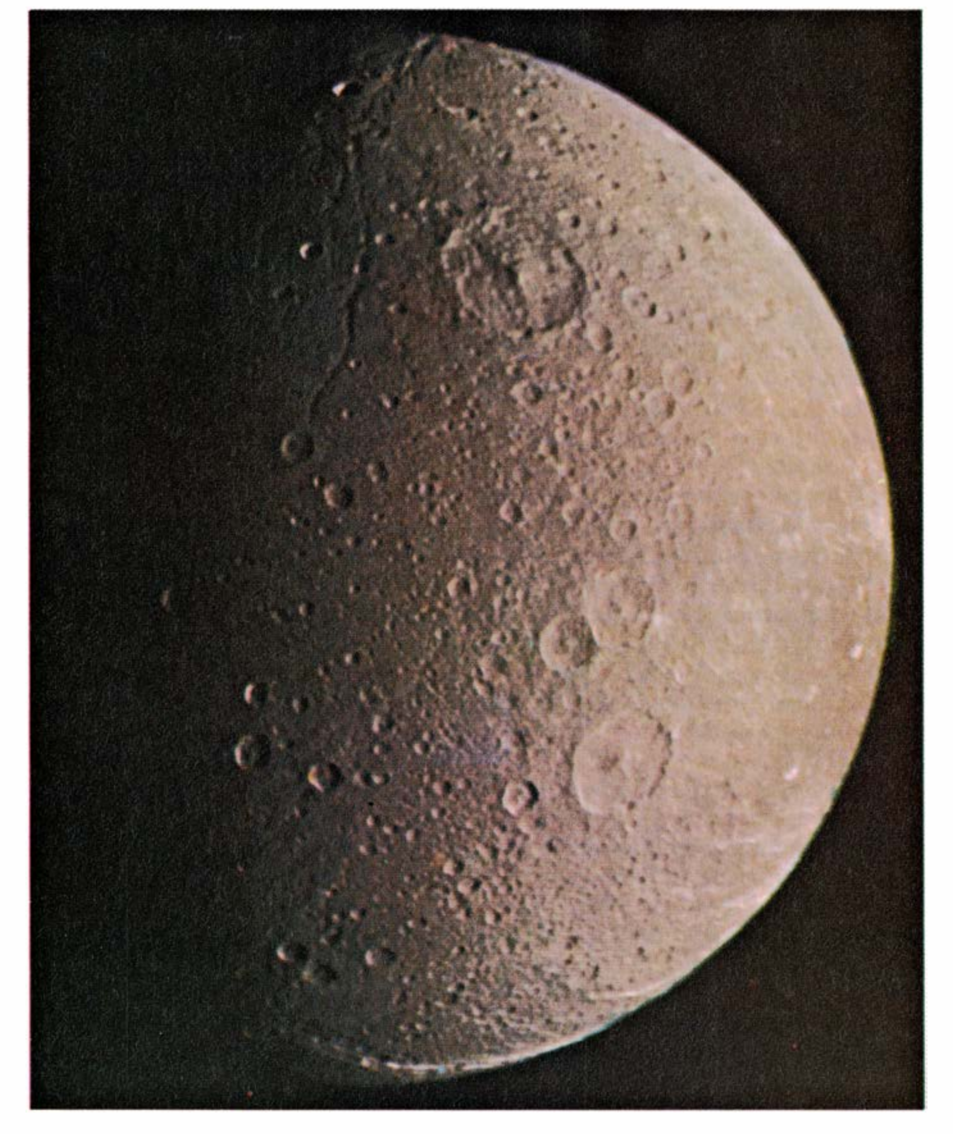
Las distantes vistas de Rhea, transmitidas a la Tierra por el *Voyager 1* cuando el vehículo espacial se acercaba al satélite, mostraron un hemisferio anterior brillante y liso, marcado sólo por lo que parecía un cráter de impacto grande y relativamente reciente. El hemisferio posterior tiene una topografía distinta. Ofrece una complicada configuración de ringleras brillantes sobre un fondo más oscuro que el hemisferio anterior. Se cree que las ringleras son el resultado de actividad interna. Tal vez, el material brillante se extravasó a lo largo de líneas de fractura de la superficie. La zona de ringleras queda confinada dentro de un pequeño círculo en el centro del hemisferio posterior. La frontera entre esta zona y el liso hemisferio anterior es difusa.

Las vistas de alta resolución de Rhea, hechas por el *Voyager 1*, muestran la región ecuatorial del hemisferio

anterior algo al Este del supuesto, y enorme, cráter reciente de impacto. Las mejores tomas recogen el polo Norte del cuerpo, sobre el que el *Voyager 1* pasó a una distancia de sólo 59.000 kilómetros. En cada una de las tomas se ve que Rhea tiene una superficie densamente salpicada de cráteres, muy semejantes a las altiplanicies con cráteres que se ven en la Luna y Mercurio. La diferencia principal reside en que los cráteres grandes y recientes de estos dos últimos están rodeados de capas de materia expulsada de ellos, mientras que los cráteres de Rhea no lo están. Presumiblemente, la responsable de ello sea la débil gravedad de Rhea. Si Copérnico, un gran cráter de la Luna terrestre, se hubiera formado en Rhea, su materia expulsada se habría distribuido sobre la mayor parte de la superficie del cuerpo. En la Luna terrestre, la porción mayor de la materia expulsada apenas trasciende unas pocas veces el radio del cráter.

La naturaleza y el origen de los cuerpos que produjeron cráteres en Rhea pueden inferirse de un mosaico de imágenes que muestran la región polar norte del cuerpo. Diríase, a primera vista, que la región se halla uniformemente salpicada de cráteres. Un examen más detenido revela que los dos tercios occidentales del mosaico polar están marcados por un conjunto de cráteres cuyos diámetros oscilan entre 30 y 100 kilómetros; hay, además, una densa población de cráteres menores, de algunos kilómetros de diámetro. De existir cráteres todavía menores, quedan por bajo del límite de resolución. El tercio oriental del mosaico está también marcado de cráteres, pero no se encuentran los grandes. Se ven en la superficie algunas depresiones poco profundas, lo que avala la hipótesis de que





DIONE fue fotografiada por el *Voyager 1*. Una imagen tomada el 11 de noviembre de 1980 (*arriba*) muestra la cara del satélite cuyo enfoque se aparta de Saturno. El fondo anaranjado es la cima de las nubes del planeta, a 377.000 kilómetros del satélite. Dione se mueve hacia la derecha; el hemisferio posterior está a la izquierda. El centro del hemisferio posterior muestra franjas brillantes en un campo oscuro. Según una hipótesis, una distribución de franjas cubrió tiempo atrás la totalidad de la superficie de Dione. Fue borrada por el continuo impacto de pequeños meteoritos a lo largo de la historia del sistema solar, excepto en el centro del hemisferio posterior, donde no son muchos los proyectiles que llegan. Una imagen tomada el 12 de noviembre de 1980 a una distancia de 162.000 kilómetros (*abajo*) ofrece la parte que se comporta como si fuera la proa de Dione. El punto central del hemisferio anterior está hacia el medio del terminador a lo largo de la izquierda de la imagen. A la derecha, la cara que mira a Saturno. Los comienzos de franjas brillantes del hemisferio posterior se insinúan hacia el borde derecho del satélite.

los cráteres grandes del Este se rellenaron o quedaron enterrados.

Esa diferencia entre las dos clases de terreno implica un período de formación de cráteres en el que diversos proyectiles con un amplio margen de energía abrieron cráteres con un amplio margen de tamaños. En algún momento, los restantes proyectiles, cuyos impactos podían haber producido grandes cráteres (de diámetros superiores a 50 kilómetros), fueron barridos. Continuó el bombardeo de objetos que podían formar cráteres menores. Mientras tanto, parte de la superficie de Rhea se regeneró, quizá por salida de líquido procedente del interior o por flujo de una mezcla de materia impulsada hasta la superficie por la presión de gases acumulados dentro del satélite. La aparición de una nueva superficie afectó el tercio oriental del terreno polar. Estimas de la velocidad a la que se están formando hoy cráteres en Rhea por cometas y asteroides indica que pocos de los visibles (grandes y pequeños) en las imágenes son recientes. Así pues, el desarrollo de tales accidentes topográficos hubo de transcurrir, en su mayor parte, en época temprana de la historia del sistema solar; debió ocurrir, además, de un modo intenso. Los proyectiles podrían haber sido restos de la acumulación de gas y polvo en órbita alrededor del Sol; quizá también, restos de la acumulación de gas y polvo en órbita alrededor de Saturno; y cabe pensar, asimismo, que llegaran en oleadas, cual restos de cuerpos, por ejemplo, que chocaron en el primitivo sistema saturniano.

Dione

La morfología de Rhea sugiere, por tanto, que dos poblaciones de proyectiles, al menos, marcaron su superficie. ¿Qué decir de Dione, el satélite que antecede a Rhea en proximidad a Saturno? La topografía del satélite permite deducciones más detalladas. Dione describe una órbita alrededor de Saturno a una distancia del centro del planeta de 6,2 R_s. Mide 1100 kilómetros de diámetro. Con una densidad de 1,5 gramos por centímetro cúbico, es el más concentrado de los satélites de Saturno con excepción de Titán. Las distantes vistas de Dione, tomadas por el *Voyager 1*, mostraron una sorprendente asimetría entre los hemisferios anterior y posterior. En éste se distinguía una red de rayas brillantes, sobre fondo negro, que se cortaban entre sí. Recordaba mucho el hemisferio posterior de Rhea. El hemisferio anterior mostraba un brillo uniforme. De nuevo, la región de complicadas variaciones en el albedo quedaba confinada a un pequeño círculo en el centro del hemisferio posterior, y, otra vez, la frontera de esta región con el hemisferio anterior, liso, era difusa.

Las tomas más cercanas del hemisferio posterior hechas por el *Voyager 1* revelaron que los cráteres de diámetros entre 50 y 100 kilómetros estaban cruzados por las rayas. Por tanto, las rayas hubieron de formarse bastante después del torrencial bombardeo que implican los cráteres. Con todo ello presente, podemos afirmar que los grandes cráteres de Dione y de ciertas zonas de Rhea (incluidos los dos tercios occidentales del mosaico polar) se parecen a los cráteres de altiplanicies que nos son familiares en la Luna, Marte y Mercurio. Llamaremos Población I a los proyectiles que produjeron esos accidentes en el sistema saturniano. La hipótesis más probable es que resultaron de la aglomeración de materia abandonada en órbita alrededor del Sol, tras la formación del sistema solar.

Las tomas próximas de Dione nos revelaron la existencia de llanuras donde la densidad de cráteres era mucho menor que en las "altiplanicies"; pero era allí mucho más alta la relación de cráteres pequeños a cráteres grandes. La proporción coincidía con la del tercio oriental del mosaico polar de Rhea. Tanto en las llanuras de Dione como en el terreno al Este del polo Norte de Rhea, la distribución de tamaños entre los cráteres se parece a la de los cráteres que rodean a los cráteres principales de la Luna. Se asemeja también a la de los cráteres terrestres que rodean los lugares donde se ha experimentado con bombas nucleares. En general, el predominio de cráteres pequeños sugiere un paisaje salpicado por fragmentos resultantes del impacto de proyectiles mayores o por trozos de cuerpos grandes desgajados en choques violentos. En el caso de Dione y Rhea, parece probable que los cráteres pequeños se produjeran por restos de cuerpos que chocaron dentro del sistema saturniano. Presumiblemente, por tanto, ciertas regiones de ambos satélites conservan una marca de formación de cráteres primitivos; en otras regiones, la superficie se recubrió de materia procedente del interior, que ocultó los cráteres más antiguos. Estas regiones recubiertas recibieron luego el impacto de un segundo bombardeo. Designaremos Población II a los cuerpos causantes de éste.

Todavía queda un problema. Las "altiplanicies" que evidencian formación de cráteres por la Población I y las llanuras que llevan el marchamo de la Población II no presentan correlación con la distribución global de albedo registrada por el *Voyager 1* a cierta distancia de Dione y Rhea. En el caso de Dione, por ejemplo, el hemisferior anterior, uniformemente brillante, y liso en las imágenes distantes, nos ofrece, a alta resolución, llanuras y altiplanicies. ¿Cuál fue, entonces, la causa de la distribución global de albedo? Cálculos realizados por Eugene M. Shoemaker, del Servicio de Inspección Geológica de los Estados Unidos, permiten cuantificar en qué medida los proyectiles contribuyen hoy a la formación de cráteres en los satélites de Saturno. Para toda, o casi toda, fuente de proyectiles, o casi toda situada fuera del sistema saturniano, el flujo de impactos varía notablemente del hemisferio anterior al posterior de cualquier satélite. Para Dione, Shoemaker calcula una variación de 10 a uno; en el caso de Rhea, la variación es de seis a uno. No importa que los proyectiles sean cometas o asteroides, o que tengan órbitas periódicas alrededor del Sol; la aceleración gravitatoria, inducida por Saturno en un proyectil incidente, supera ampliamente la de la trayectoria original que llevaba éste.

Shoemaker calcula, además, que desde el punto más alto de Dione hasta el lado del cuerpo (con respecto al movimiento orbital de Dione alrededor de Saturno), el flujo de proyectiles incidentes se divide solamente por dos. Desde el lado hasta la cola del hemisferio posterior se divide por cinco. Resulta de ello que la parte de Dione menos afectada por el continuo impacto de proyectiles será sólo una región restringida de la parte posterior del satélite.

Con la ayuda de los cálculos de Shoemaker se puede proponer una historia de Dione. En esta historia, la vida temprana de Dione estaba dominada por el impacto de cuerpos grandes, residuos probablemente del proceso de aglomeración en el sistema solar. Después, algunas zonas de Dione fueron recubiertas de nueva superficie. Mientras tanto, los choques entre los cuerpos residuales de la condensación del sistema solar en la proximidad de Saturno dieron por resultado cuerpos más pequeños. Algunos quedaron en órbitas alrededor de Saturno muy parecidas a la de Dione. Sus choques con éste salpicaron moderadamente de cráteres las llanuras de reciente creación. Por esa época se formaron fracturas en la superficie de Dione, que se llenaron con productos brillantes procedentes del interior.

Es probable que entonces toda la superficie de Dione tuviera el aspecto que

ahora sólo se aprecia en el centro del hemisferio posterior del cuerpo. Sin embargo, a lo largo de los últimos cuatro mil millones de años, Dione ha girado sincrónicamente y su superficie ha sido "cultivada" por el impacto de pequeños meteoritos procedentes de fuera del sistema saturniano. Los cráteres producidos por esos cuerpos pueden ser demasiado pequeños y estar demasiado dispersos para ser identificables en las imágenes de los Voyager. A pesar de todo, estos cráteres modificarían el aspecto de la superficie. De esta manera borraron los accidentes de la superficie, en una región que se extiende desde el punto medio del hemisferio anterior hasta bien dentro del hemisferio posterior. Esta historia explica la distribución global del albedo. Los meteoritos que llegaron bien avanzada la historia constituyen la Población III.

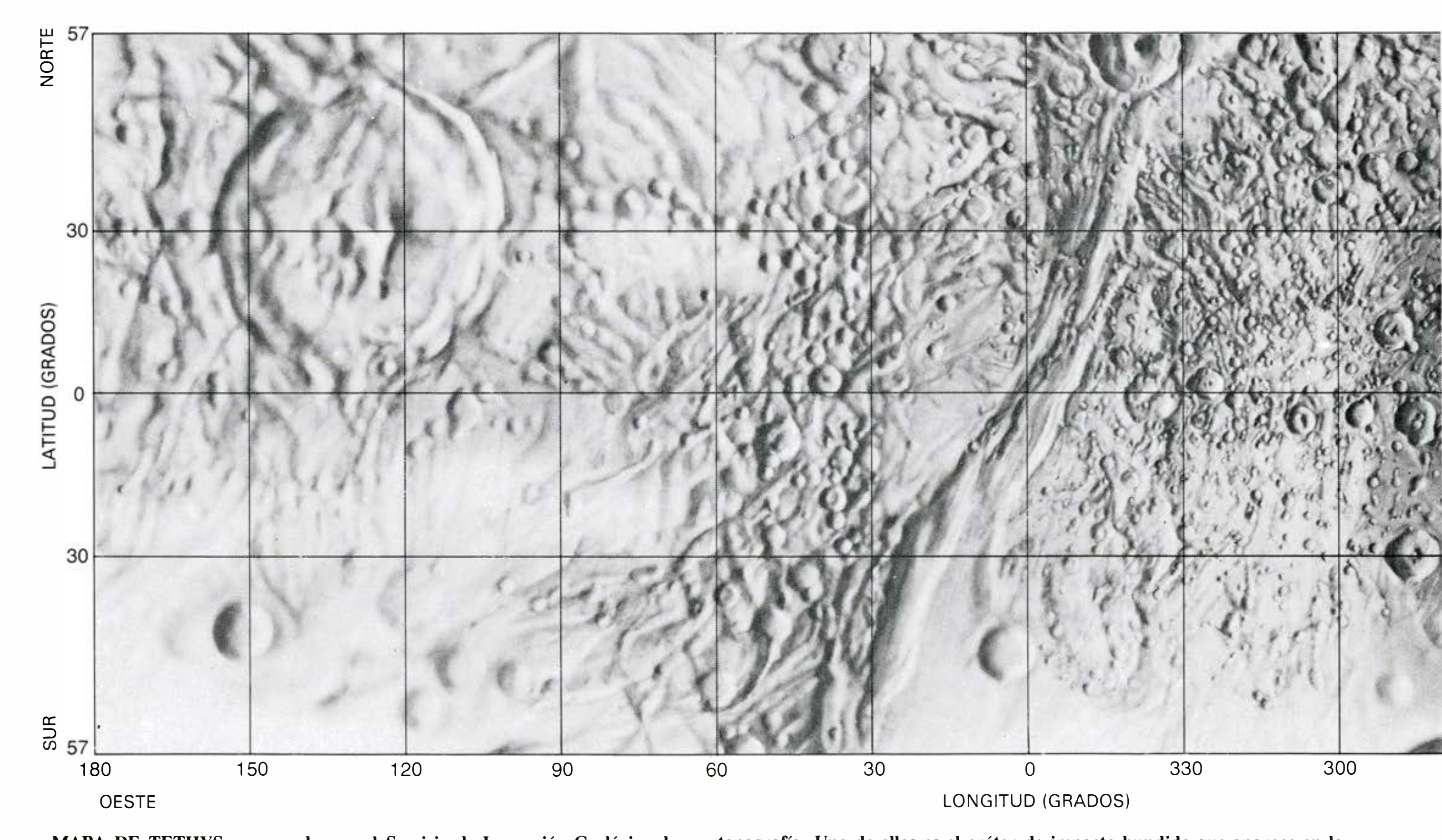
Tethys

Tras Dione, en dirección hacia el interior de Saturno, viene Tethys, que describe una órbita alrededor del planeta a una distancia de 5 R_S. Su diámetro es casi idéntico al de Dione, pero su densidad –1,2 gramos por centímetro cúbico– es más baja. Ofrece, además, una morfología totalmente distinta. Aunque el *Voyager 1* sólo pudo tomar imágenes de Tethys a baja resolución, éstas dejaban ver manchas difusas de pequeña variación de albedo en una superficie densamente cubierta de cráteres. El aspecto no se parece a la distribución global, más conspicua, que se aprecia en Rhea y Dione. En una de las mejores imágenes se distinguía un cañón bifurcado que cubría todo el intervalo que mediaba entre las regiones polares Norte y Sur del lado del satélite que mira a Saturno. Se estimó que el cañón mediría, por lo menos, 1000 kilómetros de longitud, 100 de anchura y varios kilómetros de profundidad.

Nueve meses más tarde, el *Voyager 2* se acercó a Tethys. Las imágenes que transmitió a la Tierra revelaron una enorme muesca de impacto en el hemisferio anterior. El perímetro del accidente tiene un diámetro de más de dos quintas partes del diámetro del propio Tethys. A medida que el vehículo espacial continuó acercándose, fue recogiendo fotos sucesivas cada vez más próximas a la linde visible del satélite. No tardó mucho en conseguir una imagen de su perfil. Se podía ver en ella que el fondo de lo que en su tiempo debió ser un cráter coincidía ahora con la forma esférica del cuerpo. Sólo quedaban un borde somero y un pico central rebajado. Evidentemente, el interior de Tethys estuvo en la primitiva historia del satélite lo bastante caliente para permitir el colapso de una topografía elevada. Lo mismo podría decirse de los grandes cráteres en los helados satélites jovianos Ganimedes y Calixto.

Por fin, llegó a avistarse el polo Norte y la cara opuesta a Saturno. Desde ese instante quedó claro que el cañón descubierto por el *Voyager 1* se extendía a través del polo Norte y hacia el ecuador del hemisferio opuesto a Saturno. El cañón recorría, pues, tres cuartas partes de un círculo máximo alrededor del cuerpo. Algunas partes de Tethys (como pasaba en algunas zonas de Rhea y Dione) habían recibido superficie nueva, según podía apreciarse. En una región limitada del hemisferio anterior se habían formado suaves llanuras, y el recubrimiento había borrado algunos de los grandes cráteres preexistentes.

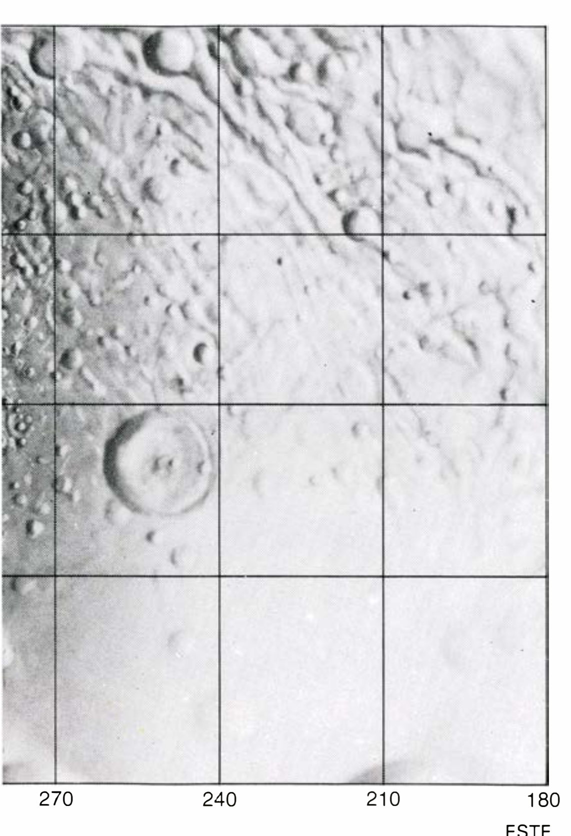
La explicación provisional de la topografía actual de Tethys arranca con la hipótesis de que, en la época en que



MAPA DE TETHYS, preparado por el Servicio de Inspección Geológica de los Estados Unidos, a partir de las imágenes hechas por el *Voyager 1* y el *Voyager 2*; es una proyección de Mercator, en la que los meridianos y paralelos son todos rectos. La longitud de cero grados indica el centro de la cara del satélite que mira a Saturno, 90 grados señala el centro del hemisferio anterior y 270 grados, el centro del hemisferio posterior. Dos accidentes dominan la topografía. Uno de ellos es el cráter de impacto hundido que aparece en la fotografía de la izquierda en la página 66. El pico central del cráter está situado a 120 grados de longitud y 30 de latitud Norte. El otro accidente corresponde al gran cañón que corre de Sur a Norte entre 30 y 330 grados de longitud. El cañón se extiende a lo largo del polo Norte de Tethys y sigue después hacia el ecuador; su extremo se ve claramente en la parte superior

se produjo la gran marca de impacto, el interior del cuerpo era mucho más móvil y estaba mucho más caliente que hoy. Tal vez era líquido. Si se hubiera tratado de un satélite frío y frágil cuando se produjo la marca, el impacto que la formó podría muy bien haberlo fragmentado. Resulta obvio, además, que las elevaciones creadas por el impacto se han derrumbado. El borde y el pico central de la huella del impacto persisten. De aquí que parezca probable que gran parte de la corteza de Tethys de la época del impacto permanezca en la superficie hoy. Podemos imaginarnos, pues, un curso histórico sin sobresaltos, en que Tethys se enfría desde la corteza hacia dentro. Si el satélite hubiera comenzado en fase líquida, la solidificación habría aumentado su volumen en aproximadamente un 10 por ciento y el área de su superficie en aproximadamente un 7 por ciento. La extensión estimada del cañón de Tethys sugiere que constituye del 5 al 10 por ciento de la superficie. El cañón puede representar el efecto de la tensión ejercida sobre la corteza al congelarse el interior y aumentar de volumen.

De los nueve satélites cuya presencia en torno a Saturno se conoció antes del siglo XX, Mimas y Enceladus son los de menor tamaño. A duras penas tienen una milésima parte de la masa de un satélite galileano de Júpiter, una milésima de la masa de la Luna y una cienmilésima de la masa de la Tierra. Su pequeñez es significativa. Antes de que los vehículos espaciales Voyager empezaran a explorar el sistema solar exterior, se creía que la principal fuente de calor de un planeta o satélite era la desintegración de los núcleos atómicos radiactivos en su interior; de ahí que pareciera que la historia de tales cuerpos viniera determinada por la relación de su volumen al área de su superficie. El volumen rige la cantidad de núcleos radiactivos y, por consiguiente, la generación de calor; el área de la superficie rige la pérdida de calor e, inversamente, su retención. La relación del volumen al área de su superficie está determinada por el tamaño del objeto: la relación aumenta con el radio. Se supone, por tanto, que los cuerpos grandes se funden con probabilidad mayor que los pequeños. La atribución del calor interno a la radiactividad, y la resultante correlación entre calor y tamaño, explican las observaciones que demuestran que la Luna está muerta, en lo concerniente a actividad volcánica, desde hace mucho tiempo, Marte posee una actividad moderada y la Tierra se muestra muy inquieta. Pero también sugiere que los satélites minúsculos, así Mimas y Enceladus, no deberían haber evolucionado substancialmente desde su época de condensación.



derecha del mapa. El cañón, que constituye de un 5 a un 10 por ciento de la superficie, puede representar la fractura de la corteza de Tethys al recubrir un interior que se dilató al enfriarse y congelarse formando hielo. En el mapa, el terreno aparece como si estuviera iluminado desde el Oeste.

Mimas y Enceladus

La imagen que de Mimas sacó el *Voyager 1* revelaba una superficie concordante con esta predicción. La superficie tenía albedo uniforme y estaba saturada de cráteres de impacto; no había sufrido modificación por vulcanismo desde los tiempos del bombardeo por la Población I. En el hemisferio anterior de Mimas hay un cráter de impacto de 130 kilómetros de diámetro, aproximadamente un tercio del diámetro del pequeño satélite. Cabe sospechar que otros cráteres del mismo tamaño, e incluso mayores, debieron producirse en el curso del bombardeo. ¿Por qué no se ven? Si un satélite frío y frágil como Mimas se hubiera fragmentado por un tal impacto, las velocidades de separación de los restos habrían sido, probablemente, del orden de la velocidad requerida para escapar del campo de gravedad del satélite. Esas velocidades hubieran sido pequeñas, comparadas con la velocidad orbital de Mimas, cifrada en 30 kilómetros por segundo. Y así, los restos hubieran permanecido en una delgada banda rodeando lo que había sido la órbita del satélite. Poco a poco, los restos se hubieran reunido de nuevo, hasta resultar en una masa fría de desechos, compuesta principalmente de hielo, con un pequeño porcentaje de roca. (La densidad de Mimas es de 1,2 gramos por centímetro cúbico.)

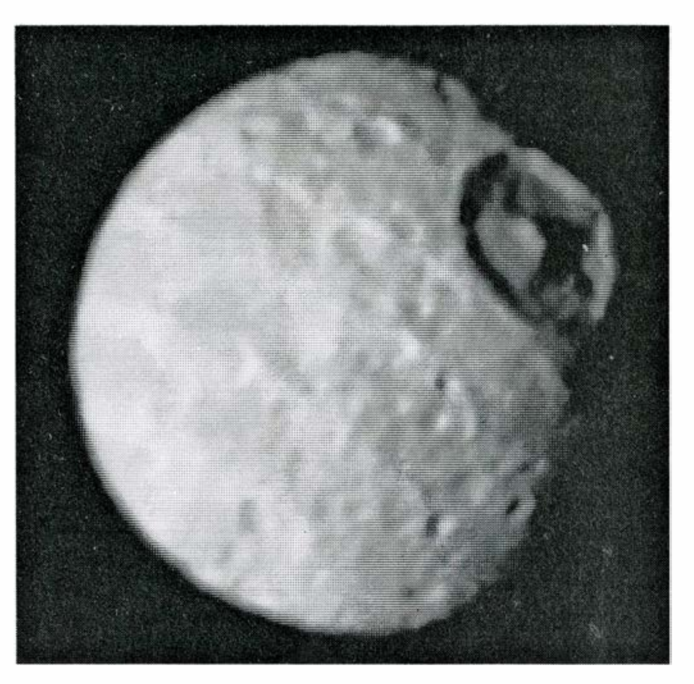
La estadística de los cráteres que se ven en Mimas sugiere que tal vez fuera eso lo que ocurrió. De los cráteres que se contemplan hoy en Mimas, mide el mayor 130 kilómetros de diámetro, para dar luego un gran salto en la escala de tamaños: los inmediatos siguientes sólo alcanzan algunas decenas de kilómetros. Mas la abundancia de cráteres aumenta exponencialmente al disminuir su magnitud. Se diría que un impacto hubiera hecho añicos el cuerpo original y que, después, se hubieran recombinado los fragmentos. La presente distribución de cráteres resulta, pues, de impactos de los últimos fragmentos. El único cráter de consideración representa un choque de energía no suficiente para haber destruido de nuevo el cuerpo; los otros cráteres grandes hechos por los objetos de la Población I desaparecieron con la fragmentación del cuerpo original.

Aunque el *Voyager 1* sólo proporcionó vistas distantes de Enceladus, los cráteres de este satélite tendrían que aparecer en las imágenes si guardaran semejanza con los de Mimas. Un cráter de la magnitud del comentado en el hemisferio anterior de Mimas debería dejarse ver, sin duda alguna. Pero las imágenes mostraban una superficie de verosímil carácter liso. Más importante, la superficie de Enceladus brillaba más que la de los satélites vecinos Mimas y Tethys. Esto sugería que gran parte de la superficie de Enceladus pudo haberse regenerado y estaba ahora cubierta de hielo de origen reciente.

Dos líneas de razonamiento habían ya indicado que Enceladus estaba fuera de lo corriente. En primer lugar, observaciones hechas desde la Tierra habían mostrado que el anillo *E* de Saturno, difuso, presentaba su brillo máximo alrededor de la órbita de Enceladus. Ello sugería que el satélite podía ser la fuente del anillo; en concreto, Enceladus podía estar continuamente reponiendo el anillo, porque, de otra manera, el campo gravitatorio del satélite tendería a eliminar materia de su órbita. (Se ha propuesto que ese pastoreo gravitatorio es responsable, cuando menos, de parte de la distribución en bandas descubierta por los Voyager en los anillos de Saturno.)

En segundo lugar, un argumento de-





EL GRAN CRATER DE IMPACTO de Tethys (*izquierda*) difiere bastante del que apreciamos en Mimas (*derecha*). Mide el primero más de 400 kilómetros de diámetro. Fue fotografiado por el *Voyager 2* a una distancia de 826.000 kilómetros. Debió tratarse de un cráter profundo, de borde y pico central abruptos. Hoy, sin embargo, se ha reconstruido el fondo del cráter, quedando a la misma altura del contorno del satélite; tanto el borde como el pico central se han hundido. Al parecer, Tethys estuvo en otra época lo suficientemente caliente para permitir tal redistribución. En esta imagen, el centro del hemisferio anterior está cerca del centro del disco del satélite. El cráter de Mimas, con 130 kilómetros de diámetro, fue fotografiado por el *Voyager 1* a una distancia de 425.000. Se trata del único ejemplar grande que posee ese satélite. Su borde y su pico central son prominentes. Al parecer, Mimas ha estado frío desde hace mucho tiempo. El pico central del cráter ocupa el centro del hemisferio anterior. (La densidad del minúsculo Mimas es de 1,2 gramos por cm³.)

sarrollado por Stanton J. Peale, de la Universidad de California en Santa Bárbara, y sus colegas concluía que Io, un satélite grande y el más interior de Júpiter, se mantenía substancialmente calentado por fuerzas de marea. Charles F. Yoder, del Laboratorio de Propulsión a Chorro del Instituto de Tecnología de California, propuso entonces que podría intervenir un mecanismo semejante en Enceladus. El período orbital de Enceladus es la mitad que el de Dione, lo que significa que uno y otro satélite se alinean con Saturno en puntos determinados de sus órbitas. La atracción gravitatoria de Dione sobre Enceladus actúa repetidamente en esos puntos, de modo que, aunque se trate de una atracción pequeña, hace que la órbita de Enceladus permanezca un tanto elíptica. En una órbita de tal clase, Enceladus se ve obligado a oscilar radialmente en el campo gravitatorio de Saturno, al igual que Io lo hace en el de Júpiter. El rozamiento interno que resulta de esta oscilación podría mantener caliente y móvil el interior de Enceladus. Los gases y fluidos que escapasen del interior podrían mantener brillante la superficie y borrar las muescas de cráteres. Allan F. Cook II, del Harvard College Observatory del Observatorio Astrofísico Smithsoniano, y Richard J. Terrile, del Laboratorio de Propulsión a Chorro, han avanzado la idea de que si Enceladus poseyera actividad geológica, los gases que escapan del interior de Enceladus y las partículas de hielo que escapan de la superficie podrían constituir la fuente que abasteciera de materia al anillo *E*.

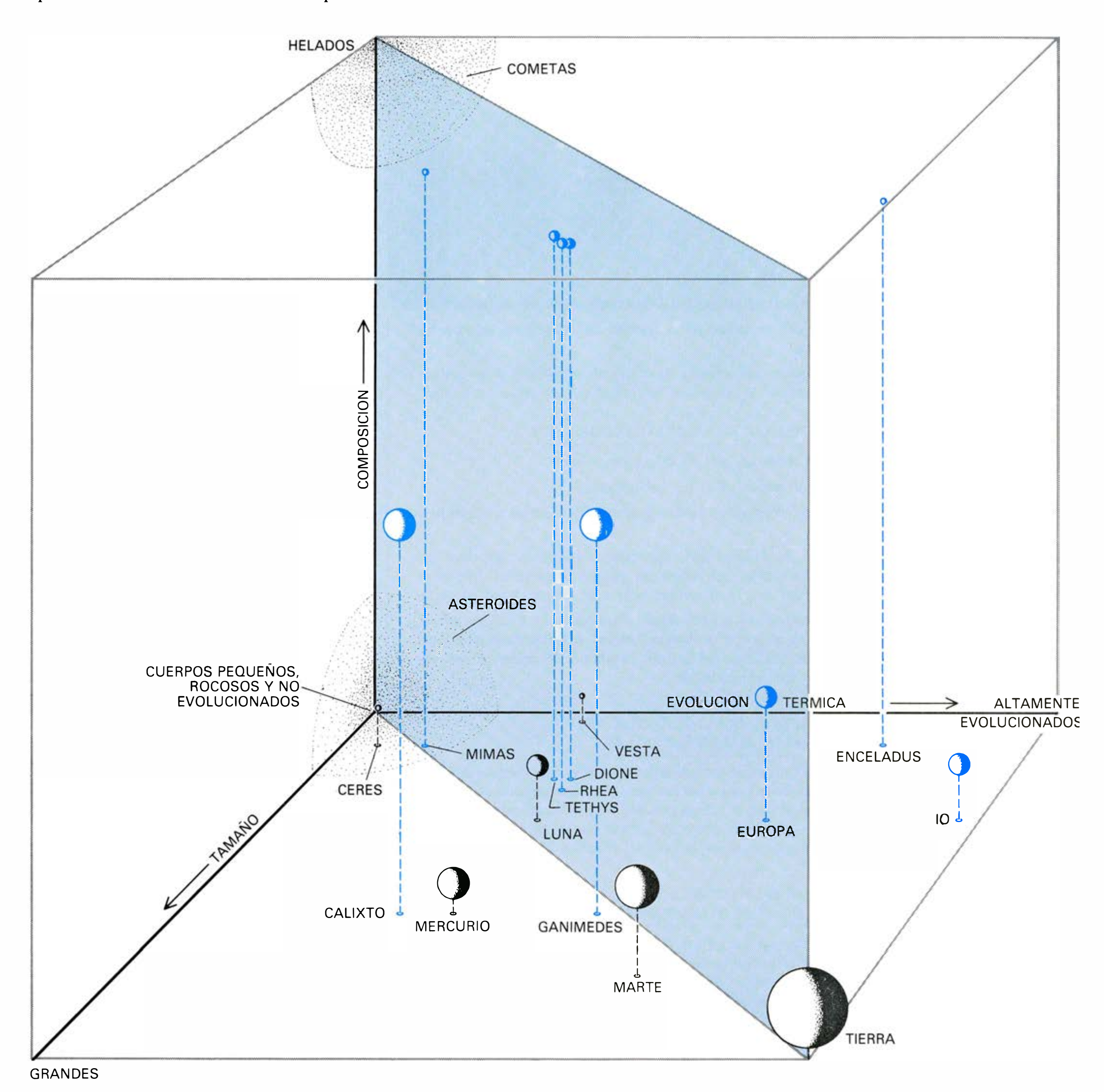
Cuando el *Voyager 2* pasó cerca de Enceladus el pasado agosto, transmitió fotos a la Tierra que, después de todo, sí dejaban ver algunos cráteres. Más escasos que los de Mimas, los hay en un estado de derrumbamiento semejante al de los cráteres viejos en Ganimedes, satélite de Júpiter. Las imágenes de Enceladus, tomadas a la máxima resolución, muestran unas llanuras definidas por una red de sierras entrelazadas, pero ninguna clase de cráteres hasta el límite de resolución. ¿Qué edad puede tener una tal superficie? Los cálculos de Shoemaker indican que los cometas que normalmente cruzan el sistema solar exterior deberían producir en las llanuras unos cráteres que habrían de observarse, a la resolución de las imágenes del Voyager, durante varios cientos de millones de años. Esto significa alrededor del diez por ciento de la edad del universo. Si otros cuerpos invisibles incidieran también sobre Enceladus, las llanuras tendrían un origen todavía más reciente.

No parece fundado admitir en Enceladus una actividad geológica a lo largo del 90 por ciento de su vida y que se haya vuelto inerte ahora. Resulta, pues, muy verosímil que Enceladus mantenga viva su actividad, de la que quizá sean consecuencia el anillo *E* y el albedo extremadamente alto de la superficie del satélite. Debe añadirse que las estimas de la efectividad del calentamiento de Enceladus por fuerzas de marea indican que la cantidad de energía suministrada por dichas fuerzas es demasiado pequeña y radiada con excesiva facilidad al espacio para que el interior sea líquido, si está compuesto de agua. Cabe que el interior incorpore clatratos de metano o hidratos de amonio. Tales sustancias tienen un punto de fusión de 100 grados C. por debajo del que caracteriza al agua pura.

Al menos seis tipos distintos de suelo podemos distinguir en la superficie de Enceladus. Los dos más antiguos, terrenos salpicados de cráteres y designados *TC1* y *TC2*, muestran una población de cráteres de 10 a 30 kilómetros de diámetro, que habrá que remontar, casi con seguridad, a un período de bombardeo de la Población II en el que Enceladus fue alcanzado por restos de materia en órbita alrededor de Saturno. En los suelos *TC1*, la mayoría de los cráteres se han derruido: sus fondos se han elevado, hundido sus bordes y sus picos centrales se han transformado en suaves cúpulas. En los *TC2*, los cráteres son profundos y en forma de tazón. De la diferencia se deduce una historia térmica desigual para esas distintas partes de Enceladus. Los terrenos *PC1* son planicies o llanuras que muestran una distribución, bastante rala, de cráteres de dos a cinco kilóme-

tros de diámetro, dispuestos a lo largo de una configuración rectilínea de fallas. Los cálculos de Shoemaker dan pie a especular sobre un intervalo de edades. Si el flujo de proyectiles recientes hubiera sido mínimo, los terrenos *PC1* podrían remontarse hasta casi la época del bombardeo por la Población II, hace más de cuatro mil millones de años. Pero si dicho flujo hubiera sido de cierta entidad, quizá tuvieran mil o incluso dos mil millones de años menos. En cualquier caso, se formaron también en Enceladus tres clases de terrenos todavía más recientes: *PC2, PC3* y *PR*. Se revelan por la presencia de desfiladeros que atraviesan los suelos más antiguos; parte de los cráteres quedan, pues, borrados.

En el *PR* se incluyen las llanuras carentes de cráteres. Su borde está defini-



CURSOS EVOLUTIVOS MUY DISTINTOS seguidos por planetas y satélites del sistema solar, según cabe deducir al representar los planetas y satélites con superficie sólida visible en un cubo cuyos ejes señalan su composición (cuerpos rocosos hacia la parte inferior y cuerpos de hielo en la parte superior), su desarrollo térmico (desde cuerpos fríos, no evolucionados, hacia la izquierda, hasta cuerpos volcánicos altamente evolucionados a la derecha) y su tamaño (de cuerpos pequeños hacia atrás a cuerpos grandes hacia delante). Las esferas negras son cuerpos cuya posición podía representarse antes de que los vehículos espaciales Voyager alcanzaran, en sus sucesivas misiones, el sistema solar exterior. En general, esos cuerpos (Mercurio, la Tierra, la Luna y Marte) son rocosos. Además, a mayor tamaño más clara evidencia de los signos de evolución y de actividad volcánica. En la ilustración, la composición rocosa de los cuerpos significa que están situados hacia la parte inferior del cubo; la correlación entre su tamaño y su evolución térmica implica que se hallan cerca del plano que atraviesa el cubo diagonalmente de izquierda a derecha. La correlación entre tamaño y evolución implica que su calor interno procede principalmente de la desintegración de núcleos de átomos radiactivos. Las esferas de color representan cuerpos cuya posición se puede situar en el cubo gracias al estudio de sus superficies sólidas en las imágenes de los Voyager. Tales cuerpos ocupan una gran parte del volumen del cubo. Incluyen Io y Enceladus, dos cuerpos pequeños y altamente evolucionados cuyo calor procede, se cree, del rozamiento interno al moverse cada uno en una órbita ligeramente elíptica alrededor de su planeta. Io es un satélite rocoso mientras que Enceladus es de hielo. Io, Europa, Ganimedes y Calixto son los cuatro grandes satélites de Júpiter descubiertos por Galileo; Ceres y Vesta son cuerpos del cinturón de asteroides, ubicado entre las órbitas de Marte y Júpiter.

do por sierras entrelazadas. Se han propuesto dos hipótesis provisionales para el origen de las sierras. De acuerdo con la primera, un sistema de fallas fue invadido por agua líquida que después se congeló. La dilatación del hielo produjo la elevación de la topografía. Sostiene la otra explicación que una zona de convección en estado sólido hace ascender la materia lentamente en el centro de la llanura y hundirse en las otras regiones, produciendo una configuración de arrugas burdamente concéntricas. Enceladus se suma a Io a la hora de demostrar que la radiactividad no siempre domina el calentamiento de los cuerpos en el sistema solar. Enceladus tiene una cienmilésima de la masa de la Tierra, lo que no le impide ejercer una actividad geológica parangonable a la de ésta.

Cursos de evolución distintos

La exploración de los sistemas de satélites que ciñen a Júpiter y a Saturno, a través de los Voyager, ha aportado nuevas ideas sobre la evolución de los cuerpos menores del sistema solar. Imaginemos un diagrama tridimensional cuyos ejes estén determinados por medidas del tamaño de un cuerpo, composición y evolución térmica. Antes de las misiones Voyager, los únicos cuerpos que podrían haber aparecido en tal diagrama hubieran sido los planetas, satélites y asteroides del sistema solar interior. Todos ellos caen esencialmente en la misma clase de objetos de idéntica composición rocosa; su grado de evolución térmica está estrechamente relacionado con su tamaño, correlación que sugiere que el calentamiento radiactivo ha dominado su historia. En la estela de las misiones Voyager, una extensa zona del diagrama gana ocupantes, deshaciéndose la tendencia, sencilla, de evolución y tamaño.

Queda mucho por descubrir acerca de la evolución de los satélites del sistema solar. Pero ya se ha esclarecido que las fuentes de energía no radiactivas (por ejemplo, calentamiento por mareas) son importantes para algunos de ellos. Además, queda demostrado que una composición a base de hielos admite una vigorosa actividad geológica, incluso en satélites muy pequeños. Los estudios que seguirán en torno a los satélites habrán de desvelarnos muchas cosas sobre las condiciones en que se hallaba y los procesos que atravesó el sistema solar incipiente, así como sobre la vía de evolución de los planetas a través de circunstancias muy varias.

Supercomputadores

El Cray-1 y el CYBER 205 ejecutan 100 millones de operaciones aritméticas por segundo. Semejante "digestión" numérica era necesaria para la solución de complejos problemas como los que se presentan en dinámica de fluidos

Ronald D. Levine

Los ordenadores de uso personal que invaden el mercado realizan varios centenares de operaciones aritméticas por segundo. En el ápice de la escala se hallan unas máquinas muy potentes: los supercomputadores. De éstos hay sólo unas tres docenas en todo el mundo. Su velocidad máxima de funcionamiento rebasa los 100 millones de operaciones por segundo. El primer ordenador electrónico comercial, el UNIVAC I comprado por la Oficina del Censo de los Estados Unidos en 1951, triplicaba la velocidad que alcanzan nuestros ordenadores de uso personal y multiplicaba por mil el tamaño de éstos. Desde la aparición del UNIVAC, la velocidad de los grandes ordenadores de uso científico se ha duplicado cada dos años, por término medio. Incremento en la celeridad que ha implicado grandes aumentos en la capacidad de la memoria necesaria para almacenar datos y resultados. Los dos últimos eslabones agregados a esta cadena son el Cray-1, construido por la Cray Research, Inc., y el CYBER 205, fabricado por la Control Data Corporation.

Aunque los actuales niveles de rendimiento de tales máquinas se deben, en gran parte, a los rápidos avances de la microelectrónica, no menos importantes han sido los nuevos conceptos introducidos en la arquitectura de ordenadores. Por arquitectura se entiende aquí la organización lógica del ordenador desde el punto de vista del programador. Las innovaciones arquitectónicas más significativas son las que permiten realizar muchas operaciones similares a la vez. Mientras que los primeros ordenadores obligaban al programador a dividir su problema de cálculo en una secuencia de pasos elementales que debían ejecutarse uno tras otro, los últimos supercomputadores le permiten especificar qué pasos elementales diferentes pueden ejecutarse simultáneamente.

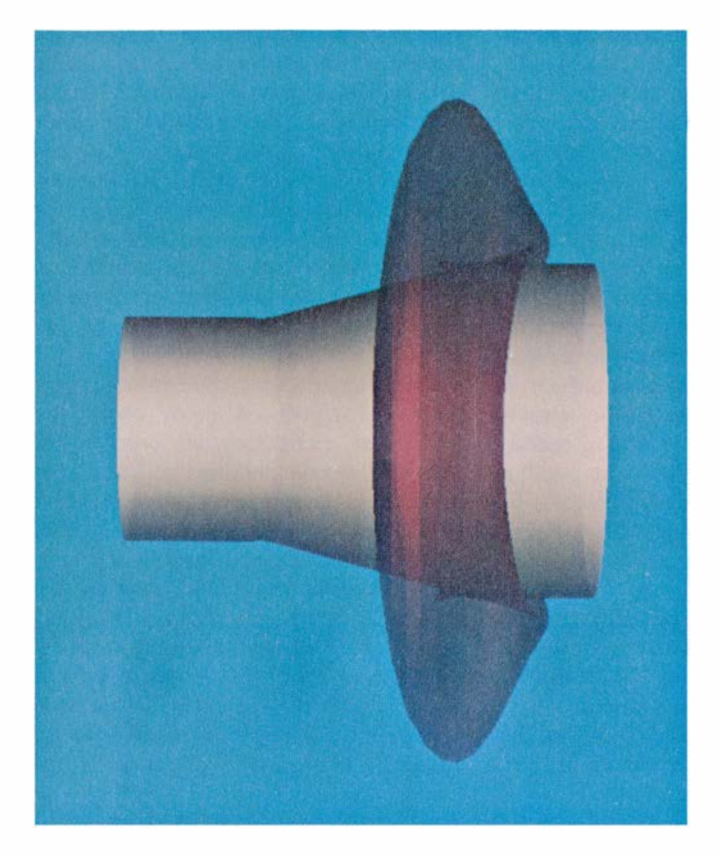
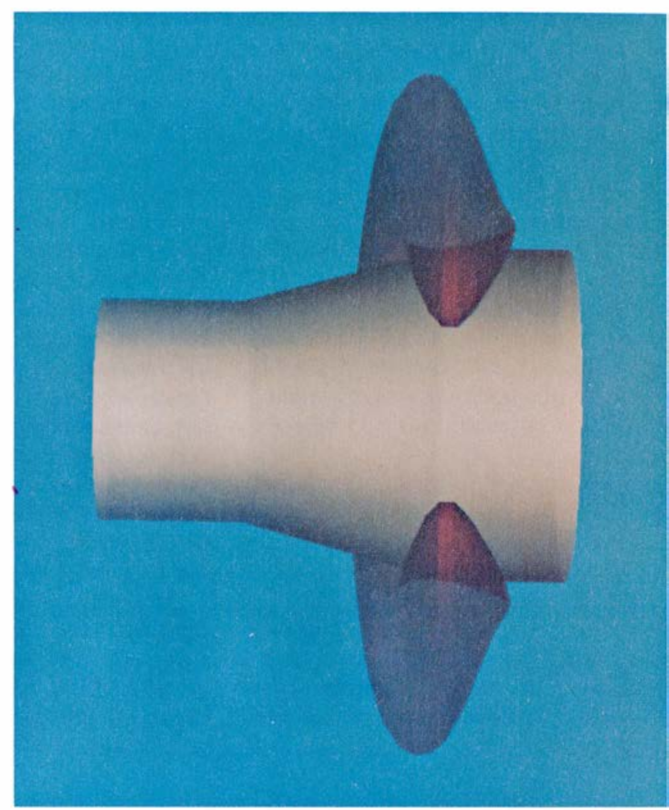
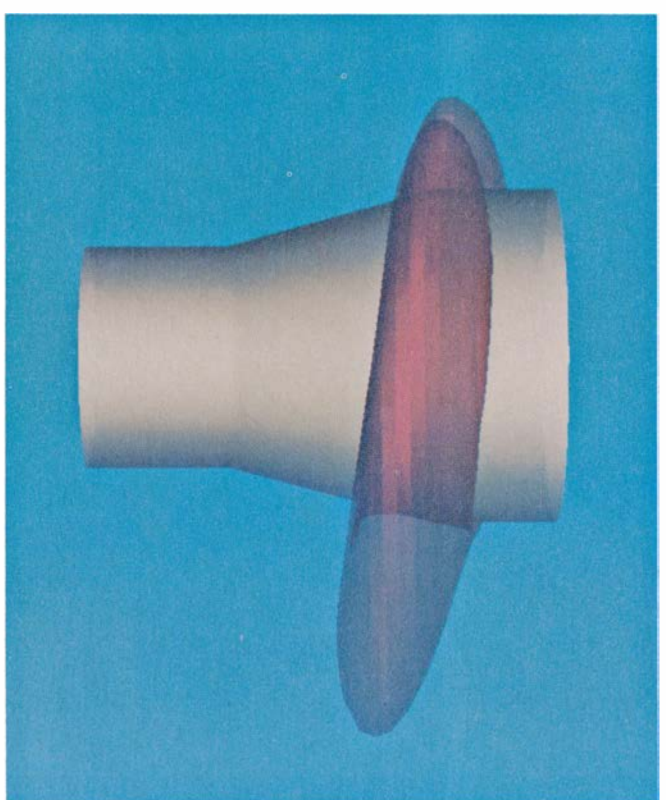
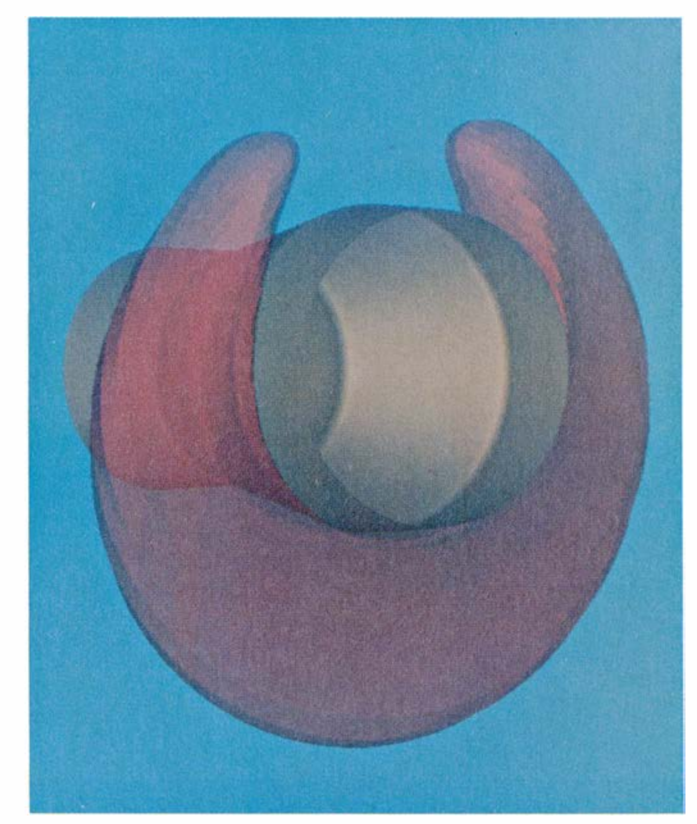
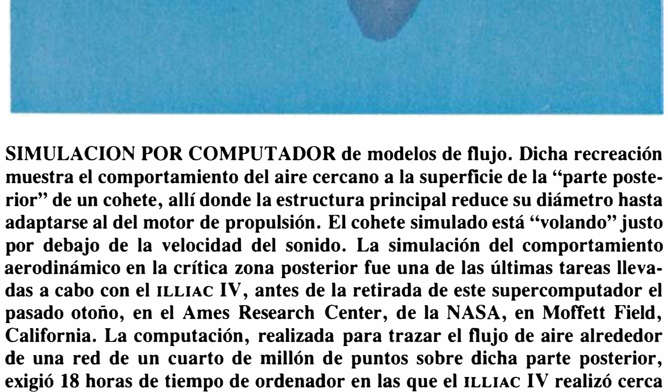
¿Qué tipo de problemas postulan la prodigiosa capacidad de cálculo que han ido adquiriendo las sucesivas generaciones de máquinas de alta velocidad? ¿Hasta qué punto permiten esos problemas emplear el procesamiento en paralelo? Bajo el techo de la simulación numérica de campos continuos se ampara una vasta categoría de problemas, excepcionalmente arduos desde el punto de vista computacional, que abarca desde la aerodinámica hasta la sismología, pasando por la meteorología y la física atómica, nuclear y de plasmas. Podemos representar un campo continuo por un conjunto de magnitudes físicas ligadas a cada punto de una determinada región del espacio tridimensional. El valor numérico de esas magnitudes puede variar de un punto a otro y cambia normalmente con el tiempo, en respuesta a alguna fuerza actuante. Las magnitudes designarán, por ejemplo, la densidad, velocidad y temperatura de un gas en movimiento, el esfuerzo mecánico de un sólido o una componente de la fuerza electromagnética.

Las leyes físicas fundamentales de los medios continuos hallan su expresión matemática en los sistemas de ecuaciones diferenciales en derivadas parciales, que relacionan (en términos de cálculo infinitesimal) los valores y las velocidades de cambio de las variables del campo en cada punto con sus valores en los puntos cercanos. Bajo la condición de especificar los valores del campo en un instante inicial o en una frontera de la región, el matemático puede demostrar que existe una única solución para el sistema de ecuaciones. Dicha solución constituye una descripción completa del comportamiento del sistema físico en cuestión. Sin embargo, salvo los casos más simples, es prácticamente imposible una expresión explícita de la solución exacta. El matemático debe recurrir, entonces, a los métodos numéricos que, aunque aproximados, pueden hacerse tan precisos como se desee aumentando el volumen de la computación aritmética.

Puesto que cada región del espacio contiene infinitos puntos, la descripción completa de un campo físico exige una cantidad infinita de datos. Quiere ello decir que el primer paso a dar, a la hora de emplear un método numérico aproximado, será ceñirse a una descripción discreta del campo mediante la introducción de una red finita de puntos distribuidos por la región. Los valores del campo corresponden exclusivamente a los puntos de la red. Lo que significa, hablando en un lenguaje matemático, sustituir el sistema de ecuaciones diferenciales en derivadas parciales por un extenso sistema de ecuaciones algebraicas ordinarias. El número necesario de puntos de la red dependerá de la dimensión espacial del modelo, la complejidad geométrica de la región y el detalle espacial buscado en el resultado.

Para simular flujos aerodinámicos turbulentos en geometrías sencillas, han sido precisas redes tridimensionales de hasta un millón de puntos y, de decenas de millones de puntos, para obtener soluciones suficientemente detalladas –exigidas por la ingeniería– en geometrías complicadas (verbigracia, la región circundante de un avión). Para cada punto de la red, deben almacenarse en memoria varios valores numéricos y actualizarse periódicamente a lo largo del proceso de cálculo: una magnitud por cada variable física del campo, varias quizá para describir la geometría y, a menudo, varios resultados intermedios. El número de magnitudes almacenadas por cada punto, en los programas de superordenadores para simulaciones tridimensionales de flujos aerodinámicos, oscila entre 5 y 30.

El proceso de cálculo se realiza a través de múltiples repeticiones de un lazo básico, en el que los nuevos valores de las variables del campo en cada punto

SIMULACION POR COMPUTADOR de modelos de flujo. Dicha recreación muestra el comportamiento del aire cercano a la superficie de la "parte posterior" de un cohete, allí donde la estructura principal reduce su diámetro hasta adaptarse al del motor de propulsión. El cohete simulado está "volando" justo por debajo de la velocidad del sonido. La simulación del comportamiento aerodinámico en la crítica zona posterior fue una de las últimas tareas llevadas a cabo con el ILLIAC IV, antes de la retirada de este supercomputador el pasado otoño, en el Ames Research Center, de la NASA, en Moffett Field, California. La computación, realizada para trazar el flujo de aire alrededor de una red de un cuarto de millón de puntos sobre dicha parte posterior, exigió 18 horas de tiempo de ordenador en las que el ILLIAC IV realizó cerca de 10^{11} operaciones aritméticas. Los resultados concuerdan casi exactamente con los que pueden obtenerse mediante un modelo del cohete en un túnel aerodinámico de pruebas. Las cuatro figuras del ordenador muestran la "superficie sonora", región en forma de herradura donde el flujo de aire alcanza la velocidad del sonido. La superficie posterior está vista por debajo (*arriba, izquierda*), por arriba (*arriba, derecha*), de lado (*abajo, izquierda*) y girada 45 grados (*abajo, derecha*). La computación fue hecha bajo la dirección de George Deiwert, del Ames Research Center. El programa para la visualización fue ideado por James H. Clark, de Stanford, y realizado por Philip Mast, de Silicon Graphics y Kelly Luetkemeyer, del Ames. En la página 84 se ven otros resultados gráficos derivados de la misma computación y de otras semejantes.

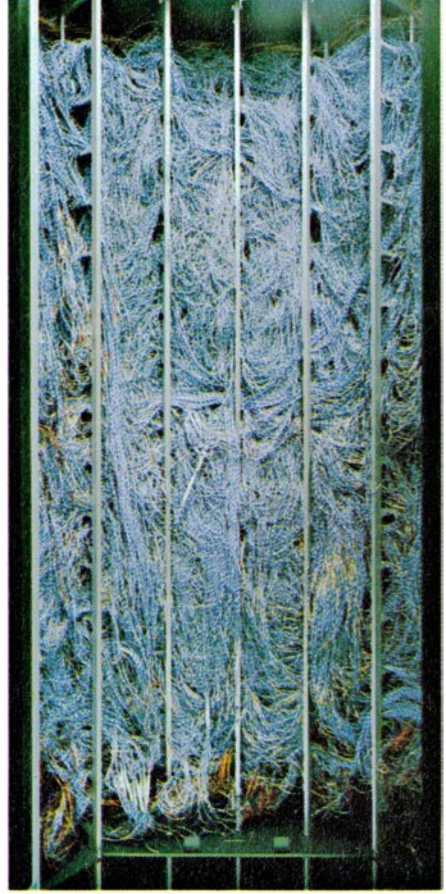
de la red se computan a partir de los valores anteriores en el mismo punto y en los vecinos. En problemas de estados estacionarios típicos, los sucesivos pasos de iteración generan aproximaciones cada vez más cercanas a la solución exacta, habiendo hecho suposiciones razonables de partida acerca de las condiciones iniciales del campo para cada punto de la red. El ordenador aplica las ecuaciones diferenciales originales al resultado de cada paso, para introducir las correcciones necesarias al objeto de que la descripción del campo se acerque cada vez más a la solución verdadera.

El número de iteraciones precisas, para lograr converger hacia una solución de suficiente aproximación a la verdadera, depende, muy directamente, del método numérico empleado y puede variar entre unos centenares y varios miles. En simulaciones donde las condiciones del campo evolucionan con el tiempo, los pasos de iteración corresponden a pequeños intervalos de tiempo. En cada paso se emplean las ecuaciones diferenciales para obtener, a partir de los valores actuales del campo, sus valores aproximados más adelante. Los intervalos de tiempo de cada paso que se empleen deberán ser menores que los precisos para que el propio campo físico sufra un cambio y, preferentemente, mucho menores si se desea obtener una exactitud numérica apropiada. A medida que los puntos de la red se aproximen, habrá que acortar los intervalos de tiempo para reducir la sensibilidad a los errores que se acumulan, inevitablemente, debido al redondeo del número de dígitos que pueden emplearse en cada paso del cálculo. El efecto final es el aumento del número de iteraciones que se precisan para llegar a un resultado útil.

Para obtener un nuevo valor de cada punto de la red, el ordenador debe ejecutar entre 10 y 500 operaciones por iteración. En algunas aplicaciones típicas de aerodinámica o física nuclear, se precisan hasta 10^{13} operaciones aritméticas, lo que lleva más de dos días de tiempo de computación. La resolución de problemas más complejos se halla pendiente de poder disponer de ordenadores más potentes. El ingeniero aeronáutico desearía obtener sus resultados de la simulación de flujos en menos de un día, en una hora si fuera posible. Otras aplicaciones, así los pronósticos del servicio meteorológico, están sometidas a restricciones más severas todavía.

En las simulaciones de campos continuos, así como en otros complejos problemas científicos, el proceso de cómputo puede dividirse en un cierto número de partes que pueden ejecutarse a la vez, aunque, por supuesto, esta posibilidad tiene sus límites. La n-ésima iteración no debe, obviamente, realizarse antes de la $n + 1$-ésima, puesto que los resultados de la primera suministran los datos de entrada de la segunda. En el caso de los problemas de campos continuos, la posibilidad de ejecución concurrente se debe a la na-



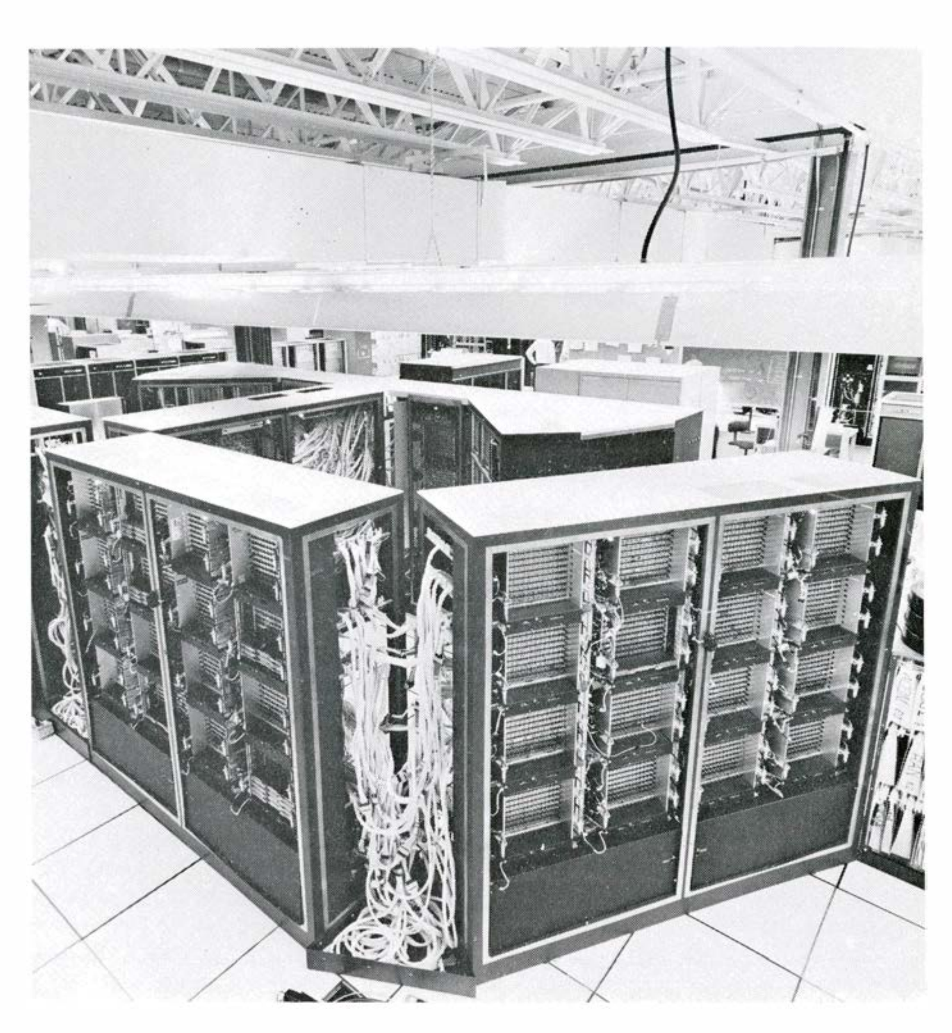


ORDENADOR CRAY-1 (*izquierda*), fabricado por la Cray Research, Inc.; es una de las dos máquinas disponibles comercialmente hoy en día capaces de alcanzar una velocidad máxima de computación de 100 millones de operaciones aritméticas por segundo. El Cray-1 mostrado en la fotografía fue recientemente instalado en el Ames Research Center, donde sustituyó al ILLIAC IV. La mayor de ambas estructuras, en primer plano, es su "bastidor principal", que contiene la unidad central de proceso y la memoria central. La estructura menor corresponde al subsistema de entrada/salida (E/S), compuesto por tres pequeños computadores rápidos y una extensión de memoria. Los bancos situados alrededor de la unidad principal alojan la fuente de alimentación. La fotografía de la derecha muestra una parte del mazo de cables de la parte posterior del Cray-1 que interconectan más de 1600 placas de circuito de la unidad central de proceso. Los cables, alrededor de 300.000 en total, tienen una longitud apropiada para que la comunicación de señales entre dos puntos cualesquiera se desvíe en menos de un nanosegundo (10^{-9} segundos) del tiempo previsto. La máquina realiza un ciclo de operación cada 12,5 nanoseg.

turaleza local de las ecuaciones diferenciales, que dotan a los puntos de la red de homogeneidad computacional. En otras palabras, si una secuencia de operaciones particular describe el cambio de las cantidades asociadas a un punto de la red entre los instantes t y $t + 1$, la misma secuencia puede ser aplicada simultáneamente a cualquier otro punto de la red, bien de toda ella, bien de los que se hallen sobre una misma sección plana. Puesto que las operaciones correspondientes a puntos diferentes son independientes, pueden realizarse a la vez. Se están dedicando muchos esfuerzos tendentes a aumentar al máximo el número de operaciones que pueden realizarse concurrentemente con la nueva generación de supercomputadores, mediante un replanteamiento de los algoritmos y el empleo de nuevos lenguajes de programación.

Suele tomarse como medida de la velocidad de los supercomputadores el "megaflop": un millón de operaciones en coma flotante por segundo. La expresión "coma flotante" se refiere a la versión binaria del conocido método de notación científica para la representación de los números. En notación científica normal, un número se representa mediante el producto de dos factores, el primero de ellos (la mantisa) posee una magnitud comprendida entre 0,1 y 1, y el segundo (la característica) es una potencia entera de 10. Así, el número 6600 se representa por $0{,}66 \times 10^4$, y el 66 por $0{,}66 \times 10^2$. Se precisa tal representación en cálculo científico debido a que, a menudo, el rango de las magnitudes del problema es muy grande. El ordenador trabaja con una versión binaria de la notación en coma flotante en la que la característica es una potencia entera de 2 y ambos términos (mantisa y característica) están expresados en notación binaria, consistente en una serie de ceros y unos.

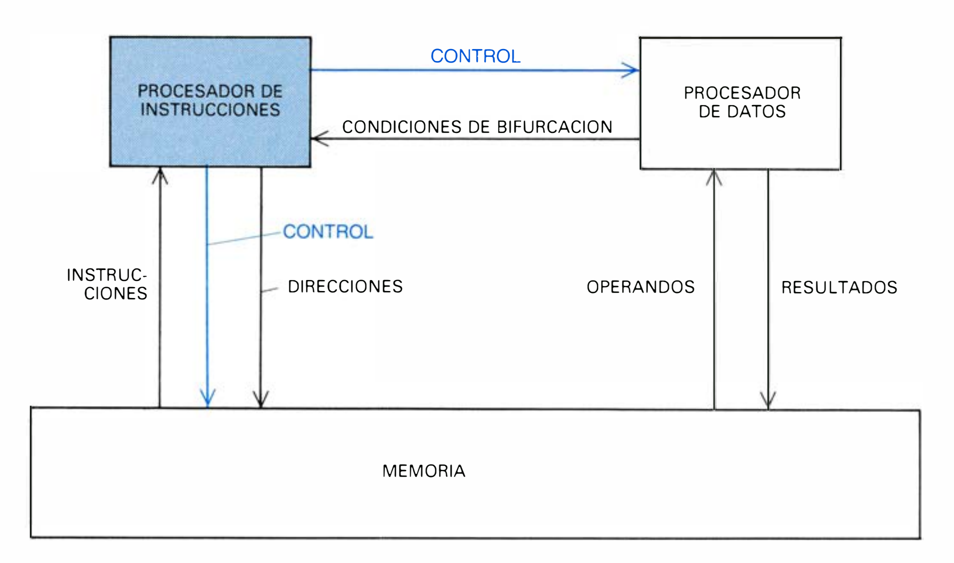
Operaciones simples en coma flotante son la suma, resta, multiplicación y división de dos operandos expresados en coma flotante para obtener un resultado expresado también en coma flotante. Tales operaciones exigen algo más de trabajo de cómputo que las correspondientes de coma fija, o enteros. En los modernos computadores, las tres primeras operaciones absorben aproximadamente el mismo tiempo, mientras que la división es algo más lenta. Sin embargo, los programas pueden escribirse generalmente de forma que el número de divisiones sea muy bajo. A la hora de calcular la velocidad en megaflops, al tiempo necesario para ejecutar las operaciones debe añadírsele el empleado en obtener los operandos desde memoria y devolver a ésta los resultados.



EN EL CYBER 205 el Cray-1 tiene un fuerte competidor, fabricado por la Control Data Corporation. La unidad que aquí se muestra está atravesando las pruebas finales en la fábrica de la compañía en Arden Hills, Minnesota. El precio del CYBER 205 y del Cray-1 oscila entre 10 y 15 millones de dólares, según la cantidad de memoria rápida y otras varias opciones que se deseen. El primer CYBER 205 fue entregado, la primavera pasada, al Servicio Meteorológico de la Gran Bretaña. El pronóstico del tiempo ha supuesto uno de los principales incentivos para el desarrollo de los supercomputadores, que se inició con el MANIAC I (analizador matemático, integrador numérico y computador) diseñado a finales de la década de 1940 por John Von Neuman y sus colegas. El CYBER 205, desarrollado a partir del STAR 100, de la C.D., tiene un tiempo de ciclo de 20 nanosegundos. Dada la gran capacidad del Cray-1, las señales pueden ir de un punto a otro a través de cables normales de cobre, a una velocidad aproximada de 0,3 veces la de la luz. Como el CYBER 205 es considerablemente mayor, los diseñadores han debido recurrir al empleo de cables coaxiales, para unir sus varios miles de placas, en los que las señales circulan a 0,9 veces la velocidad de la luz.

La evaluación en megaflops debe ser matizada indicando la precisión con que se llevan a cabo las operaciones, esto es, el número máximo de dígitos significativos determinados por la longitud de palabra de la máquina. Los supercomputadores están diseñados para trabajar con palabras largas, típicamente de 64 bits (dígitos binarios). En una palabra de 64 bits, normalmente se asignan 49 a la mantisa y su signo (más o menos) y los 15 restantes al exponente y su signo. De esta manera, pueden representarse números comprendidos entre 10^{-2466} y 10^{2466} con una precisión equivalente a 15 dígitos decimales.

Podría dudarse de la necesidad de tamaña precisión, sobre todo teniendo en cuenta que las medidas científicas son casi siempre de una precisión menor en varios órdenes de magnitud. Sin embargo, en muchos algoritmos los inevitables pequeños errores que derivan de los redondeos pueden acumularse al cabo de centenares de miles de pasos iterativos, hasta formar un error inaceptable en el resultado final, salvo que en cada operación se realice con gran precisión. Por otra parte, hay muchos algoritmos insensibles frente a los errores de redondeo y que, por tanto, pueden trabajar sin hilar tan fino. Por esta razón, una característica deseable en un supercomputador de utilización general es la posibilidad de opción a modos de trabajo de precisión media, que resuelva los problemas a doble velocidad de



LOS ELEMENTOS BASICOS EN UN COMPUTADOR son dos procesadores, que realizan activamente operaciones aritméticas y lógicas, y una unidad de memoria, que almacena pasivamente las instrucciones y los resultados de los continuos cómputos. Al principio, la memoria está cargada con las instrucciones detalladas que constituyen el programa a ejecutar. El procesador de instrucciones las obtiene en una secuencia programada y las envía al procesador de datos. Los resultados son llevados a la memoria (*abajo, en la ilustración*) de donde son extraídos para posteriores cálculos hasta terminar el programa.

la que supone cada megaflop con el doble de memoria operativa.

No puede asignarse una determinada velocidad o número de megaflops a un ordenador en particular sin tener en cuenta el programa que vaya a ejecutar. La velocidad depende de la tasa de concurrencia inherente al problema, de cuán adecuada sea su concurrencia con la capacidad de procesamiento en paralelo de que está dotada la máquina y, especialmente, de la experiencia del programador. Niveles de rendimiento, suministrados por los constructores, por encima de los 100 megaflops corresponden siempre a valores máximos, alcanzables sólo en pequeñas subrutinas y que raramente se dan en la práctica. El número medio de megaflops que puede mantenerse a lo largo de la ejecución de un programa típico desciende, invariablemente, por debajo de los máximos indicados y depende, muy directamente, del tipo de programa.

A la altura de 1982, consideraremos supercomputador, de acuerdo con una norma extensamente aceptada, a la máquina que se muestre capaz de mantener valores medios de 20 megaflops en problemas típicos de iteración que empleen una base de datos de un millón o más de palabras. La computadora de mayor éxito de la década de 1970, la Control Data CDC 7600, podía alcanzar como máximo los 5 megaflops, y eso sólo con problemas mucho más sencillos que los citados. El primer ordenador capaz de alcanzar los 20 megaflops fue el ILLIAC IV, diseñado bajo la dirección de D. L. Slotnick en la Universidad de Illinois, financiado por la Agencia de Investigación en Proyectos Avanzados (ARPA) del Departamento de Defensa y construido, a lo largo de los últimos años de la década de 1960 y primeros de la de 1970, por la Burroughs Corporation. En 1972 se instaló el ILLIAC IV en el Centro de Investigaciones Ames de la NASA, donde permitió resolver con éxito varios de los más complicados problemas aerodinámicos. Causó baja en el servicio en septiembre pasado.

Tras el desmantelamiento del ILLIAC IV, sólo quedan a la venta dos familias de supercomputadores en condiciones operativas: la Cray-1, de la Cray Research, y la CYBER 205, de la Control Data. El primer Cray-1 fue instalado en Los Alamos Scientific Laboratory, en 1976. Hay ya 34 unidades en diversos centros. El primer CYBER 205 totalmente equipado se entregó al Servicio Meteorológico de Gran Bretaña, la primavera de 1981. El resto ha sido adquirido por varios gobiernos, laboratorios universitarios y dos compañías petrolíferas. El CYBER 205 supone una evolución del STAR 100 de la Control Data, del que se construyeron 4 unidades a principios de la década de 1970. Entre ambos, apareció un modelo intermedio de transición, el CYBER 203, del que se fabricaron tres unidades en 1981, hasta la aparición del 205, y que seguramente serán mejorados para equipararlos a éste, por lo que no merece la pena detenerse en las diferencias existentes entre ambos.

No puede decirse que Cray-1 o CYBER 205 sean indiscutiblemente más rápidos que ILLIAC IV. Los tres están (o estaban) especialmente indicados para tipos particulares de problemas científicos. Dado que el ILLIAC IV constituyó un diseño único por encargo, el coste de su proyecto y construcción fue muy elevado, cerca de 40 millones de dólares; ello explica que resultara el más caro, con mucho. Tiene también un coste de mantenimiento muy alto: unos 2 millones de dólares por año. Según la memoria empleada, los periféricos y las posibles opciones deseadas, los precios de los otros oscilan entre los 10 y los 15 millones de dólares. Se advierte una fuerte competencia entre las empresas constructoras por acaparar el mercado mundial de supercomputadores, estimado en unas 100 o 200 máquinas para los próximos 5 años. Además, ambas compañías proyectan fabricar nuevos ordenadores con superior capacidad en dicho período. Para 1985, si no antes, aparecerán, cabe presumirlo, nuevos supercomputadores fabricados en laboratorios estatales y privados de los Estados Unidos y del Japón que superen a la actual generación de Cray y Control Data.

El vertiginoso aumento de las velocidades de funcionamiento de los computadores, acaecido en los últimos años, debióse, en gran parte, a la continua reducción del tamaño de los circuitos integrados microelectrónicos. El número de transistores que pueden construirse sobre una oblea de silicio, de algo más de dos centímetros de lado, ha pasado de la docena en 1960 a varios miles en 1970, para llegar actualmente a centenares de miles. Tales avances de la tecnología de semiconductores permitieron, por primera vez, construir grandes unidades de memoria de alta velocidad con un coste razonable. El ILLIAC IV fue el primer macroordenador cuya memoria central constaba de circuitos integrados, en vez de núcleos de ferrita dispuestos sobre una matriz y recorridos por finísimos conductores. Los circuitos de memoria del ILLIAC IV eran del tipo que hoy llamamos de integración en media escala (MSI): 256 bits por circuito. La capacidad total de la memoria era pequeña comparada con las actuales: 131.072 palabras de 64 bits. Las memorias centrales del Cray-1 y del CYBER 205 emplean integración en gran escala (LSI) que permite 4096 bits por circuito, con lo que ambas máquinas disponen de 4.194.304 palabras de 64 bits de memoria rápida, soportada por tecnología bipolar de silicio de tipo rápido denominada lógica de emisores acoplados (ECL) y con un coste aproximado de un dólar por palabra.

Otra ventaja del incremento de la densidad de tales circuitos concierne a la reducción del tiempo de ejecución de

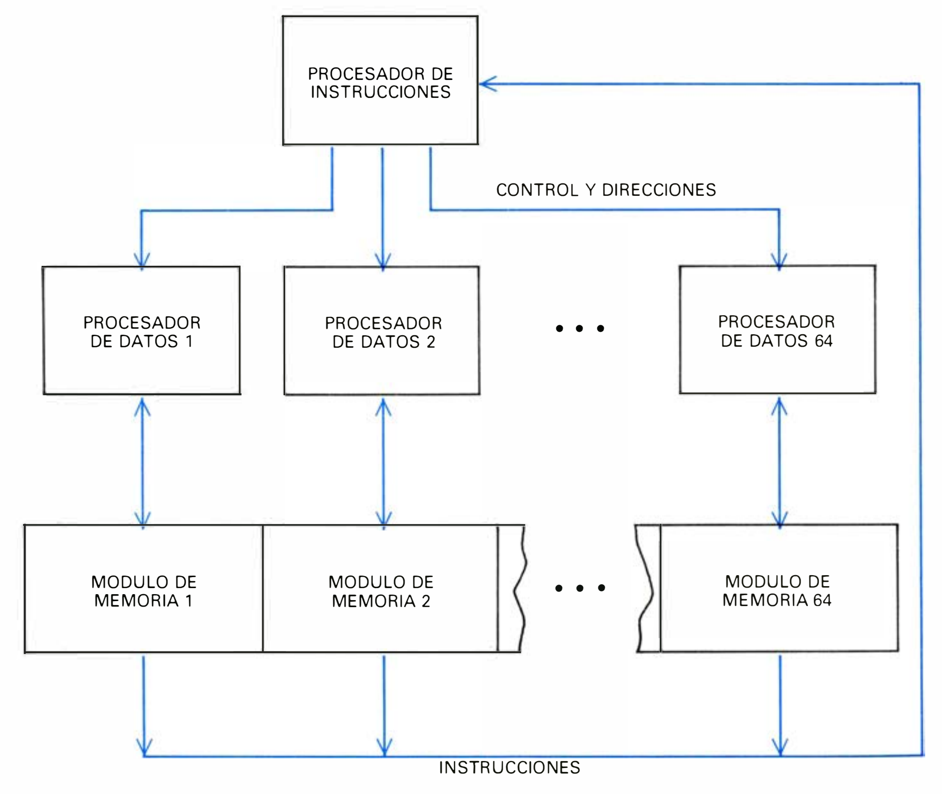
los ciclos de cada operación lógica. Los computadores digitales funcionan con intervalos temporales discretos: uno o varios generadores emiten pulsos con períodos constantes, que reciben la denominación de tiempo de ciclo o período de reloj. El estado de la máquina, entendido como el contenido de todos sus registros en un preciso instante, está determinado con exactitud en cada pulso de reloj. El cambio de estado entre pulsos depende de las entradas a la máquina. En términos digitales, nada ocurre en un tiempo menor de un período de reloj, por lo que el diseñador trata siempre de reducirlo al mínimo.

La velocidad del reloj está limitada por el tiempo de respuesta de las puertas lógicas, o conmutadores elementales, que conforman el computador. Durante muchos años, el factor limitativo más importante de los tiempos de ciclo fue el impuesto por los retardos de las puertas. En la primera tecnología de válvulas de vacío, dichos tiempos eran del orden de 10.000 manosegundos (10^{-5}segundos). Con la llegada de los transistores se redujeron hasta cerca de 100 nanosegundos. La tecnología de los circuitos integrados ha permitido descender hasta el nanosegundo (10^{-9} segundos). Con tecnologías que aún se hallan en desarrollo pueden conseguirse tiempos muy inferiores al nanosegundo, pero su empleo deberá esperar las próximas generaciones de ordenadores.

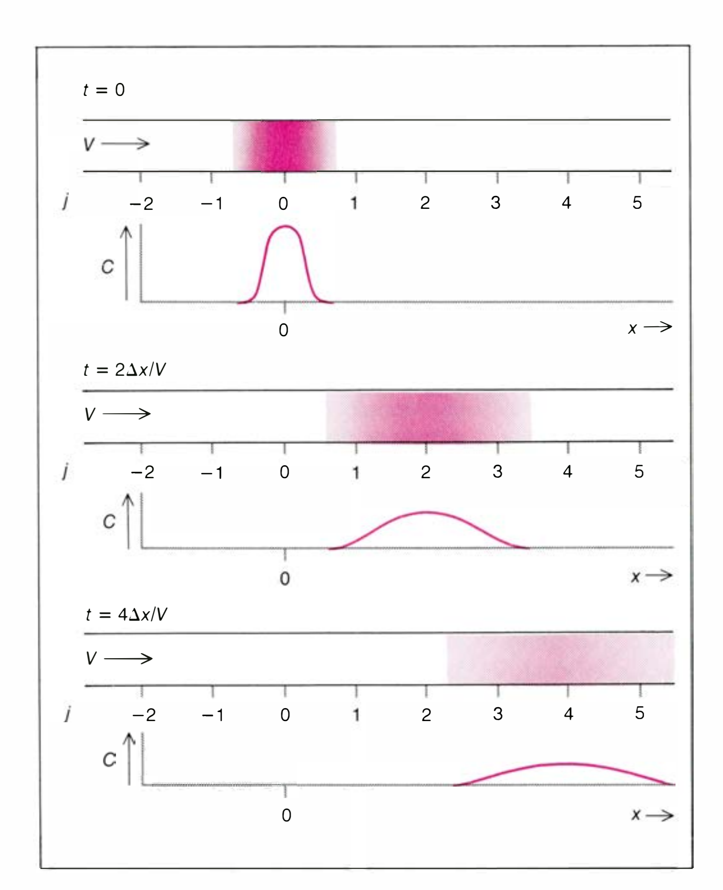
Cuando los retardos de puerta se reducen al nanosegundo, aparece una nueva limitación a la frecuencia del reloj: la velocidad de propagación de las señales desde un punto de la máquina hasta otro. Ninguna señal puede circular a mayor velocidad que la luz: 300 milímetros por nanosegundo. En la práctica, la velocidad de circulación de los pulsos eléctricos por los conductores de los supercomputadores varía entre 90 y 270 milímetros por nanosegundo. Por tanto, para reducir el tiempo de ciclo a un nanosegundo será necesario limitar la distancia que separa a las distintas partes que funcionen síncronamente a bastante menos de 300 milímetros. Aunque el incremento de la densidad de componentes de los circuitos integrados ha permitido construir dispositivos asombrosamente pequeños, así los microprocesadores (computadores prácticamente completos en una pastilla de circuito integrado), la miniaturización de los supercomputadores está limitada por la gran cantidad de memoria y circuitería lógica que deben poseer. El tiempo de ciclo del ILLIAC IV era de 80 nanosegundos; los del CYBER 205 y del Cray-1 son de 20 y 12,5 nanosegundos, respectivamente. Dichos tiempos guardan una razón proporcional inversa con las dimensiones físicas de la máquina, lo que constituye una prueba de que la velocidad de propagación es el factor limitante.

La concentración de muchos circuitos en un espacio reducido, para minimizar la longitud de los cables, trae consigo un serio problema: la eliminación del calor generado por las conversiones de energía eléctrica. El consumo de los elementos lógicos rápidos bipolares es considerable: unos 5 watt por pastilla; el Cray-1 contiene cerca de 300.000 pastillas en un volumen de menos de 2,7 metros cúbicos. La eliminación del calor resulta imperiosa, puesto que la tasa de fallos de los dispositivos semiconductores aumenta rápidamente con la temperatura. El ILLIAC IV era refrigerado por grandes volúmenes de aire frío que circulaban por sus chasis. En el Cray-1 y el CYBER 205 se elimina el calor mediante freón comprimido, que circula por tubos de refrigeración. En el Cray-1, las placas de circuito, que poseen un tamaño de 15 por 22 centímetros y contienen docenas de pastillas de circuitos integrados, están unidas a sólidas chapas de cobre cuyos bordes de ambas caras se hallan en estrecho contacto con grandes placas verticales de aluminio, por las que circula el refrigerante. En el CYBER 205 los tubos refrigerantes quedan directamente embutidos en las placas de circuito impreso; cada pastilla se mantiene en íntimo contacto con uno de los tubos mediante un muelle de presión.

Los circuitos digitales están sujetos a errores provocados por fallos de los componentes y por ruido eléctrico. Aunque los circuitos integrados son cada vez más fiables, las tasas de errores crecen a medida que evolucionan los supercomputadores; ello obedece a que en los sistemas complejos dichas tasas de error son, como poco, proporcionales al número de componentes. Para proteger al usuario frente resultados erróneos, los supercomputadores incluyen códigos matemáticos capaces de detectar y corregir los fallos. ¿Cómo? Introduciendo bits de información redundante a la longitud mínima de palabra. Las técnicas matemáticas originalmente ideadas en 1950 por Richard W. Hamming, de los Bell Laboratories, mostraron de qué modo añadir, con eficacia, los bits redundantes. Por ejemplo, añadiendo ocho bits a una palabra de 64 cabe detectar los errores que aparezcan en dos bits a la vez, y detectar y corregir los que se provoquen en uno.



ORGANIZACION DEL ILLIAC IV. Se trata de un supercomputador único en su género diseñado en el Center for Advanced Computation de la Universidad de Illinois, a finales de la década de 1960, que comprendía 64 procesadores de datos independientes, cada uno con su propio módulo de memoria, bajo el control de un único procesador de instrucciones. La máquina podía realizar 64 computaciones afines simultáneamente. Hay muchos problemas en meteorología, aerodinámica y otras disciplinas que pueden resolverse con tal estructura. El ILLIAC IV aún no ha sido superado a la hora de resolver los problemas más complejos. Fue el primer gran ordenador en usar los circuitos de memoria con semiconductores.



a

TERMINO DE CONVECCION — TERMINO DE DIFUSION

$$\frac{\partial C}{\partial t} + V\frac{\partial C}{\partial x} + D\frac{\partial^2 C}{\partial x^2} = 0$$

DERIVADA TEMPORAL — VELOCIDAD DEL FLUIDO — GRADIENTE DE CONCENTRACION — CONSTANTE DE DIFUSION — DERIVADA SEGUNDA DE LA CONCENTRACION

b

$x \rightarrow j\Delta x \qquad j = 0, \pm 1, \pm 2, \ldots$

$t \rightarrow n\Delta t \qquad n = 0, 1, 2, 3, \ldots$

$C \rightarrow C_j^n$

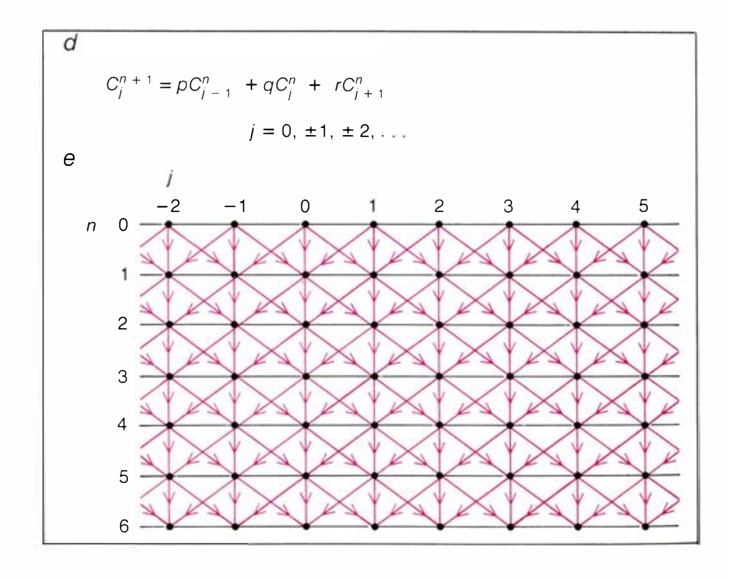
c

$$\frac{\partial C}{\partial t} \rightarrow \frac{C_j^{n+1} - C_j^n}{\Delta t}$$

$$\frac{\partial C}{\partial x} \rightarrow \frac{C_j^n - C_{j-1}^n}{\Delta x}$$

$$\frac{\partial^2 C}{\partial x^2} \rightarrow \frac{\dfrac{C_{j+1}^n - C_j^n}{\Delta x} - \dfrac{C_j^n - C_{j-1}^n}{\Delta x}}{\Delta x}$$

UN PROBLEMA TRATABLE por procesamiento paralelo en un supercomputador es el relativo al flujo de fluidos (*izquierda*): en un canal largo y estrecho se mueve un fluido, con velocidad constante V hacia la derecha, que contiene una gota de líquido contaminante (*color*). El contaminante es arrastrado por la corriente (convección) y también difunde debido al movimiento molecular al azar (difusión). Un gráfico situado debajo de cada figura muestra la concentración C del contaminante en función de la distancia x a lo largo del canal. El problema matemático consiste en escribir las ecuaciones que darán el valor de C en cada posición x del canal para cada instante t. Los puntos de la red, denotados por j y separados una distancia $\triangle x$, se emplean para la descripción discreta del problema en la simulación por computador. El valor $\triangle x$ debe ser lo suficientemente pequeño para poder conseguir la resolución espacial deseada. La dinámica del problema del flujo de fluidos puede describirse por una ecuación en derivadas parciales conocida como ecuación de transporte (a, *a la derecha*). C, la concentración de contaminante, varía de un punto a otro y en cada uno de ellos a lo largo del tiempo. El primer término de la ecuación corresponde a la variación temporal, el segundo al efecto de convección; V es la velocidad de la corriente y su derivada parcial es el gradiente de concentración, esto es, la rapidez de cambio con respecto a su posición a lo largo del tubo. El tercer término ofrece el efecto de la difusión. D es una constante que refleja las propiedades difusivas de ambos fluidos; la derivada parcial segunda es el gradiente del gradiente de concentración. En *b*, el planteamiento continuo ha sido sustituido por una aproximación discreta. Se ha tomado un pequeño intervalo de tiempo $\triangle t$; sólo se considera la concentración en los instantes $t = 0$, $\triangle t$, $2\triangle t$, ..., y exclusivamente en los puntos j. El término C_j^n denota el valor de la concentración en el punto j de la red en el intervalo n. Deben tomarse límites a izquierda y derecha para j en orden a imponer la finitud para que el problema sea tratable por computador. Aquí basta con restringir el cómputo a las regiones en que la concentración se anula prácticamente. El número de cálculos necesarios para cubrir la evolución del sistema a lo largo del período de interés depende del valor de $\triangle t$. Para mantener una cierta precisión, $\triangle t$ debe ser pequeño aun a costa de alargar la computación. En *c*, cada derivada parcial ha sido sustituida por su "mejor" aproximación, tal como calculará en los puntos específicos de la red espacio-temporal. En cada caso, tal aproximación es una fracción cuyo denominador es el intervalo espacial de la red y cuyo numerador es la diferencia entre los valores de los puntos adyacentes. En el caso de la derivada segunda, el numerador es la diferencia entre elementos que son, a su vez, fracciones de ese tipo.



MISION DEL ORDENADOR es resolver un amplio sistema de ecuaciones, *d*, obtenido al hacer la sustitución por la aproximación discreta de *c*, dada en la figura superior de esta misma página, en la ecuación de transporte. Tras la simplificación algebraica se obtiene una ecuación para cada valor de j que ofrece la concentración en ese punto j en el instante $n + 1$ en función de la concentración en el instante n anterior en el punto j y en sus dos vecinos, $j - 1$ y $j + 1$. Los factores p, q y r son constantes que dependen de V, D, $\triangle x$ y $\triangle t$, e independientes de j y n. Para obtener la solución en el siguiente intervalo de tiempo, el ordenador debe realizar cinco operaciones en coma flotante por cada punto: tres multiplicaciones y dos sumas (el término "coma flotante" se refiere a la notación científica en que los números se expresan como potencias de 10). El diagrama (*e*) de la red espacio-temporal indica las interdependencias entre todos sus puntos. Antes de calcular el dato correspondiente a un punto, se requieren como entradas los asociados a los puntos de los que parten las flechas coloreadas hacia aquél (se supone conocida la concentración inicial, $n = 0$, para todo j). Todos los valores de j para cada n pueden, en principio, ser computados en paralelo, o concurrentemente, puesto que no hay interdependencias entre los puntos de un mismo nivel temporal. Tales valores pueden calcularse una vez que se hayan deducido los correspondientes al instante anterior. Por ser la estructura de cómputo igual para todos los puntos j, puede programarse la computación en paralelo para procesamiento vectorial.

Deben incluirse los circuitos de generación y comprobación de los bits de redundancia en varios puntos de control de la máquina. Todos los supercomputadores de la actual generación incluyen dichos esquemas de control de errores. Su ausencia en el ILLIAC IV, junto con un soporte físico ("hardware") propenso a fallos, se convirtió en verdadera pesadilla. La seguridad que proporciona el control de errores incrementa considerablemente el costo de la máquina y no elimina la posibilidad de errores no detectables (de tres bits o más), aunque reduce su tasa de aparición a niveles aceptables.

Antes de entrar en la discusión de la arquitectura de los ordenadores, permítaseme hacer un breve repaso de las funciones básicas que realizan. Un computador consta de tres partes esenciales: la memoria, el procesador de instrucciones y el procesador de datos (por simplicidad omito el sistema de entrada/salida encargado de la comunicación con el medio exterior). La memoria contiene los datos y las instrucciones; sus posiciones son referidas mediante direcciones únicas. Por dato se entiende cualquier elemento que pueda ser tratado como operando, es decir, que pueda manipularse matemáticamente, lo que incluye a los resultados elaborados con anterioridad. Las instrucciones son patrones de bits codificados adecuadamente que especifican las operaciones a realizar y las direcciones de memoria a las que haya que accede. Las instrucciones se agrupan en programas, o códigos.

El conjunto de instrucciones, u operaciones de soporte físico definidas, constituye la esencia de la arquitectura. El procesador de instrucciones las obtiene de la memoria, las decodifica y envía las adecuadas secuencias de señales de control y de direcciones al procesador de datos y a la memoria, de forma que se ejecuten las operaciones indicadas por el código de la instrucción. Lo más frecuente es que la siguiente instrucción a procesar sea la posterior a la presente en la memoria. Son pocos, sin embargo, los programas construidos por simples secuencias lineales de instrucciones; la mayoría contienen lazos (secuencias de instrucciones que deben repetirse muchas veces) y saltos condicionales, en los que la próxima instrucción a ejecutar depende de dos o más posibilidades según el resultado que se obtenga en el programa, y que no puede ser predicho.

En los computadores convencionales de tipo secuencial, cada instrucción puede abarcar sólo un pequeño número de operandos; las arquitecturas más comunes permiten, desde luego, una única dirección de memoria por instrucción. Obtener tal instrucción de la memoria y decodificarla por el procesador de instrucciones supone emplear varios ciclos de máquina en ello. Para ejecutar la instrucción deben emplearse varios ciclos en el acceso a memoria y varios más para realizar las operaciones aritméticas. En los primeros ordenadores, el procesador de instrucciones entraba en estados inactivos de espera durante la fase de ejecución de la instrucción. Pronto resultó evidente el ahorro de tiempo que podía suponer que el procesador de instrucciones recabase de memoria la próxima instrucción, mientras el de datos trabajaba con la presente, siempre y cuando no se produjera un conflicto de concurrencia a memoria entre ambos. Este solapamiento entre búsqueda y ejecución fue el precursor de los evolucionados conceptos llamados de telescopado ("pipeline") que hoy dominan el diseño de los computadores vectoriales.

Para el programador, un vector es una lista ordenada de datos cuyos elementos deben almacenarse en memoria de cierta forma regular. El número de tales elementos de la lista recibe el nombre de "longitud del vector" (definición que difiere de la más familiar que se da en física y en matemáticas). Para cada operación que haya de aplicarse a un único operando (tal como la extracción de la raíz cuadrada) existe su correspondiente operación vectorial, consistente en aplicarla a cada elemento del vector. De la misma forma, para una operación a aplicar a dos operandos hay su correspondiente operación vectorial, que debe aplicarse al par de elementos tomados ordenadamente de los respectivos vectores, que serán de igual longitud. Para distinguirlas de las vectoriales, las operaciones aritméticas ordinarias realizadas sobre números únicos o pares de números se llaman operaciones escalares.

Los computadores vectoriales, superordenadores incluidos, cuentan con varias estrategias que permiten acelerar la ejecución de las operaciones vectoriales. Una de ellas consiste en la inclusión de instrucciones vectoriales como parte del conjunto total de las órdenes. La ejecución de una instrucción vectorial se traduce en la ejecución del correspondiente par de operaciones equivalentes al total. Así, además de indicar la operación a efectuar, la instrucción vectorial especifica las direcciones de los dos operandos vectoriales, el vector resultado y su longitud. Más significativo es el hecho de que el soporte físico ("hardware") permite la ejecución simultánea de muchas operaciones de dos operandos que constituyen la operación vectorial. Tal tipo de concurrencias pertenecen, por regla general, a alguna de estas dos clases: multiproceso y telescopado.

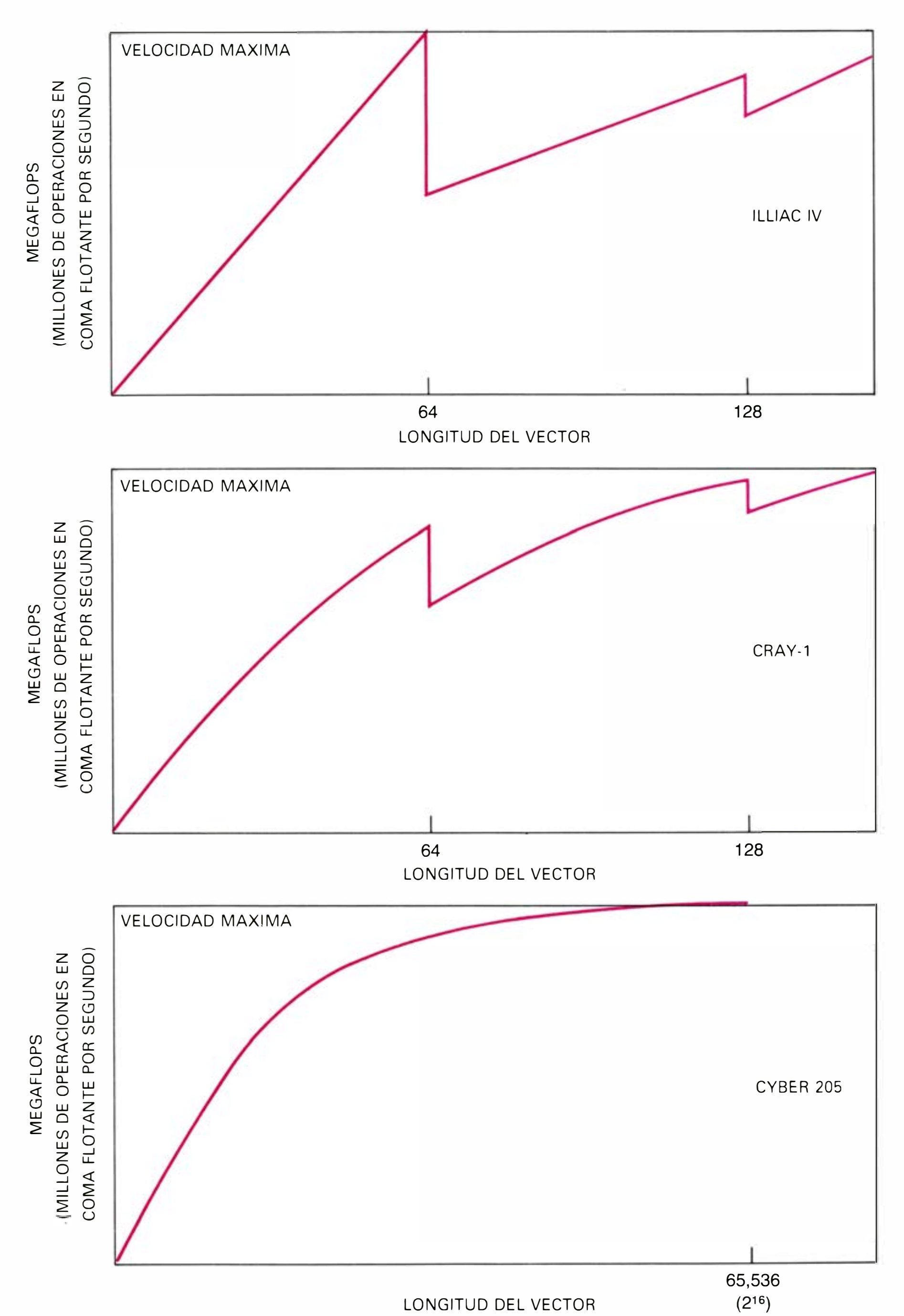
El ILLIAC IV incluía un forma primitiva de multiproceso. Esta máquina poseía 64 procesadores idénticos de datos que operaban en pasos concatenados bajo el control de un único procesador de instrucciones. La memoria estaba subdividida en 64 módulos, cada uno de los cuales dedicado a un procesador de datos. Los elementos del vector ocupaban las correspondientes posiciones en los módulos de memoria. De acuerdo con esto, la longitud vectorial óptima era 64, de forma que las operaciones se realizaban con una velocidad 64 veces mayor que en el caso escalar ordinario. Para el caso de operaciones de mayor longitud, se dividían en segmentos de 64, que se ejecutaban en un lazo del programa a imagen de lo que ocurre en toda máquina secuencial. Para vectores de dimensión menor de 64 (o para la parte última de un vector mayor no divisible por 64) podían detenerse algunos procesadores; en ambos casos, el resultado era una disminución del número de megaflops. El ILLIAC IV podía funcionar como 128 procesadores cuando la longitud de palabra pudiera acortarse hasta 32 bits sin que ello significara una pérdida apreciable de precisión.

El telescopado puede asimilarse a una cadena de montaje industrial donde el producto en fabricación atraviesa una serie de estaciones. En cada una se efectúa una operación particular y todas ellas trabajan simultáneamente con diferentes unidades en distinto estadio de elaboración. Por ejemplo, la suma en coma flotante comprende varios pasos que deben ejecutarse secuencialmente. Un sumador en coma flotante telescopado se divide en segmentos, cada uno de los cuales realiza una parte del trabajo en un período de reloj. Al final de este período, cada segmento comunica sus resultados al siguiente y recibe los del anterior. Así, al cabo de un tiempo de duración igual a tantos períodos de reloj cuantos segmentos posea el sumador, se obtendrá a la salida la suma en coma flotante correspondiente a los valores que en un principio se le suministraron. Tales datos deben ofrecerse en cada período de reloj, y su suma estará disponible a la salida también durante el mismo lapso de tiempo.

Las arquitecturas del Cray-1 y del CYBER 205 emplean con profusión la técnica del telescopado interpolada con ciertas etapas de multiproceso adicionales.

Durante años se ha venido dedicando mucha atención a la organización de la memoria. Hay dos formas básicas de subdividirla y que podríamos llamar vertical y horizontal. La razón de la estructuración vertical, también conocida como jerarquizada, reside en el hecho de que las memorias rápidas tienen un mayor coste por bit que las lentas. Además, cuanto mayor es la memoria, más tiempo se precisa para acceder a los datos almacenados con anterioridad. En la mayoría de los macroordenadores, las unidades de procesamiento se comunican directamente con una memoria pequeña, muy rápida, con capacidad para unos centenares de palabras, accesibles en un ciclo de máquina. Denominada usualmente conjunto de registros, esa memoria se considera parte constitutiva de la propia unidad de procesamiento.

El siguiente nivel de la jerarquía lo forma una memoria grande y rápida, conocida como memoria central o principal. En ella reside la parte principal del programa y la base de datos durante la ejecución de los extensos problemas de iteración. Algunos diseños incluyen una memoria "fantasma", capaz de contener algunos miles de palabras, entre la principal y los registros.



PROCESAMIENTO VECTORIAL, una forma de paralelismo computacional disponible en los modernos superordenadores. Un vector es, simplemente, una lista ordenada de datos que pueden manipularse en paralelo con los datos de otra lista de igual longitud. El soporte físico ("hardware") de los computadores está preparado para la ejecución simultánea de largas ristras de tales operaciones. Debido a las diferencias de su arquitectura, la longitud vectorial óptima varía de uno a otro supercomputador (ILLIAC IV, Cray-1 y CYBER 205), representados por las tres curvas. Se llama "megaflop" al millón de operaciones aritméticas en coma flotante por segundo. La velocidad máxima de cada máquina depende de la operación vectorial en particular. Para el CYBER 205 (*figura inferior*) puede llegar hasta 400 megaflops.

En el nivel inferior de la jerarquía se encuentra un sistema de almacenamiento secundario, relativamente lento pero de gran capacidad, frecuentemente bajo la forma de un conjunto de discos magnéticos. El Cray-1 y el CYBER 205 tienen la misma unidad de discos de alto rendimiento, construida por Control Data, capaz de almacenar 77 millones de palabras. A ambas máquinas se les puede acoplar una batería de tales unidades, con lo que llegan a superarse los mil millones de palabras de 64 bits. Los datos pueden transferirse hasta, o desde, tales unidades de disco a una velocidad máxima de medio millón de palabras por segundo; en la práctica, puede realizarse un intercambio simultáneo de datos entre la unidad central y varias unidades de discos.

Se denomina ancho de banda a la máxima velocidad de transferencia de información hasta, o desde, los dispositivos de memoria. Para que el ancho de banda de los niveles inferiores de la jerarquía de memoria no limite la velocidad media de cómputo, deben organizarse los programas de forma que el mayor número posible, tanto de instrucciones como de datos, se hallen en los niveles superiores para cada cálculo sin precisarse, durante su ejecución, la descarga de información desde un nivel inferior hasta otro superior. Como veremos más adelante, esta consideración afecta, especialmente, al Cray-1, cuya memoria central posee un ancho de banda pequeño en relación con la tasa de megaflops que puede alcanzarse con los datos contenidos en los registros.

El almacenamiento secundario en discos del Cray-1 y del CYBER 205, demasiado lento, no permite una velocidad razonable en la simulación de campos continuos cuya base de datos de iteración es excesiva para acomodarse en la memoria central. El ILLIAC IV tenía un sistema exclusivo de memoria de disco que disponía de una cabeza de lectura/escritura para cada pista. Todos los discos trabajan sincronizados; un buen programador podía llegar a conseguir velocidades de transferencia de 7 millones de palabras de 64 bits por se-

gundo. Como consecuencia, el ILLIAC IV, a pesar de su primitiva arquitectura y pequeña memoria central, se convirtió en el computador más rápido en la solución de los grandes problemas jamás construido. Resolvía simulaciones aerodinámicas que involucraban hasta 8 millones de palabras de 64 bits o 16 millones de palabras de 16 bits de la base de datos iterativos. La Control Data y la Cray Research, así como varios fabricantes más, acaban de poner a punto una nueva memoria situada en un nivel intermedio entre la central y la de discos. Este nivel estaría realizado con tecnología MOS (metal-óxido-semiconductor) y podría almacenar hasta 32 millones de palabras con un costo de 10 centavos de dólar por palabra. Los primeros Cray-1 o CYBER 205 equipados con dicha memoria tomarían del ILLIAC IV la calificación de "computador más rápido del mundo".

Lo descrito hasta ahora es la estructura horizontal de la memoria del ILLIAC IV, cuya memoria central fue dividida en 64 módulos que trabajan en paralelo, dedicado cada uno a un procesador. Varios supercomputadores con multiproceso, en fase de desarrollo, incorporan un refinamiento de este esquema: cierto número de módulos de memoria paralela independientes están ligados a otros tantos procesadores independientes, a través de un conmutador de alta velocidad bajo control de programa, de forma que todas las memorias son accesibles desde todos los procesadores.

Para los procesadores telescopados, se piensa ya en otra clase de distribución horizontal de la memoria central: la memoria está dividida en un determinado número de bancos de memoria "desfasados", así llamados porque trabajan de forma que sus respectivos ciclos de acceso se hallan desplazados en fase unos de otros. La razón fundamental para esta distribución es que las memorias centrales, de acceso aleatorio, son relativamente lentas, siendo preciso un número mínimo de períodos de reloj entre sucesivas referencias. Para mantener la secuencia de operandos vectoriales a una velocidad de una palabra por período de reloj, como exige la estructura telescopada, los vectores se almacenan con los consecutivos operandos en diferentes bancos. El desplazamiento de fase que "abre" los bancos referenciados sucesivamente coincide con un período de reloj del procesador. El CYBER 205 tiene 16 bancos, el Cray-1 puede manejar 8 o 16, según el tamaño de la memoria. En ambas máquinas, el tiempo de un ciclo de memoria corresponde a 4 ciclos del procesador. Así, el ancho de banda de la memoria central del Cray-1 es de una palabra por ciclo de 12,5 nanosegundos, vale decir, 80 millones de palabras por segundo. En el CYBER 205, la distribución en bancos desfasados se combina con una partición cruzada para formar un conjunto de módulos paralelos llamados pilas (stacks): 8 pilas para una memoria central de un millón de palabras y 16 pilas para memorias con 2 a 4 millones de palabras. El máximo ancho de banda para esta memoria de 16 pilas es de 16 palabras por período de reloj de 20 nanosegundos, 800 millones de palabras por segundo.

La unidad central de proceso del CYBER 205 puede disponer de dos o de cuatro sistemas telescopados para realizar operaciones aritméticas en coma flotante. El modelo que dispone de cuatro debe poseer, por lo menos, una memoria central de 2 millones de palabras que impliquen el ancho de banda requerido para mantener tales sistemas a pleno rendimiento. Todos ellos pueden sumar, restar, multiplicar, dividir y

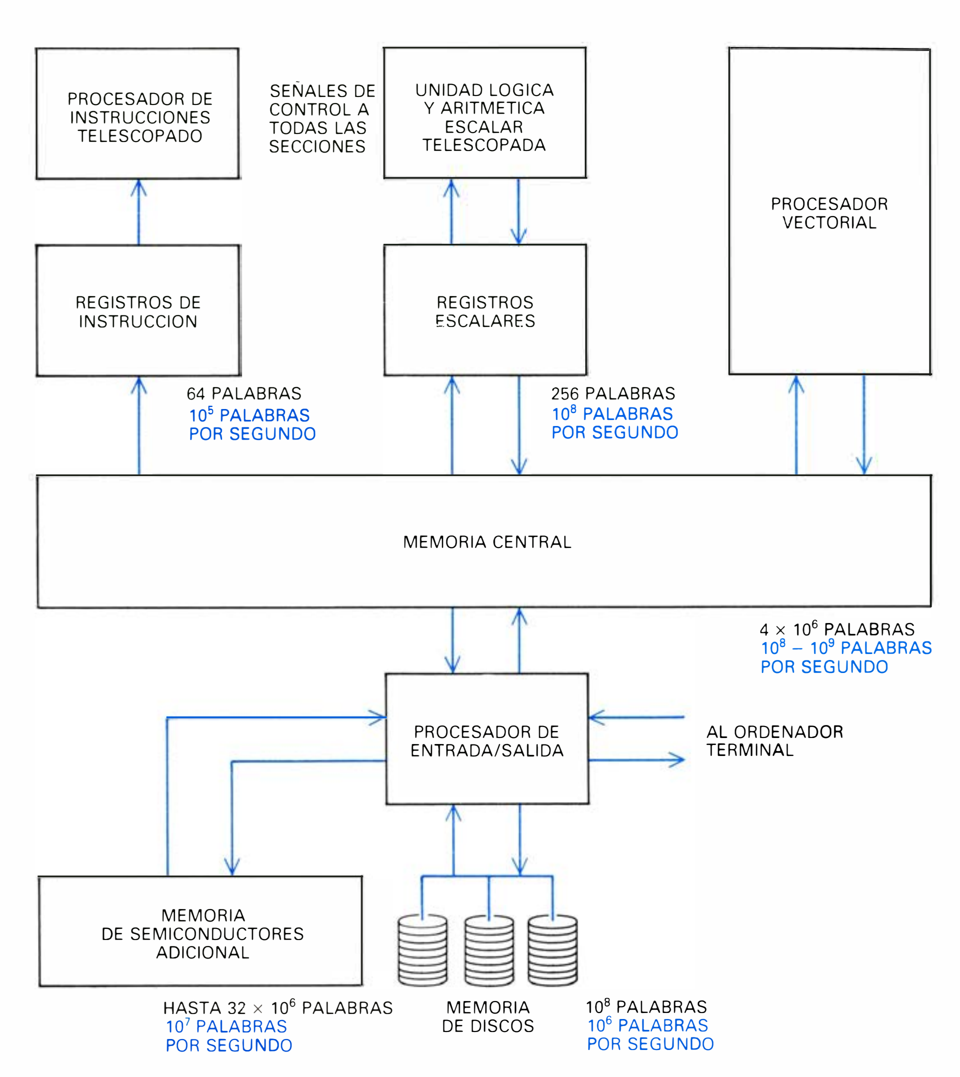
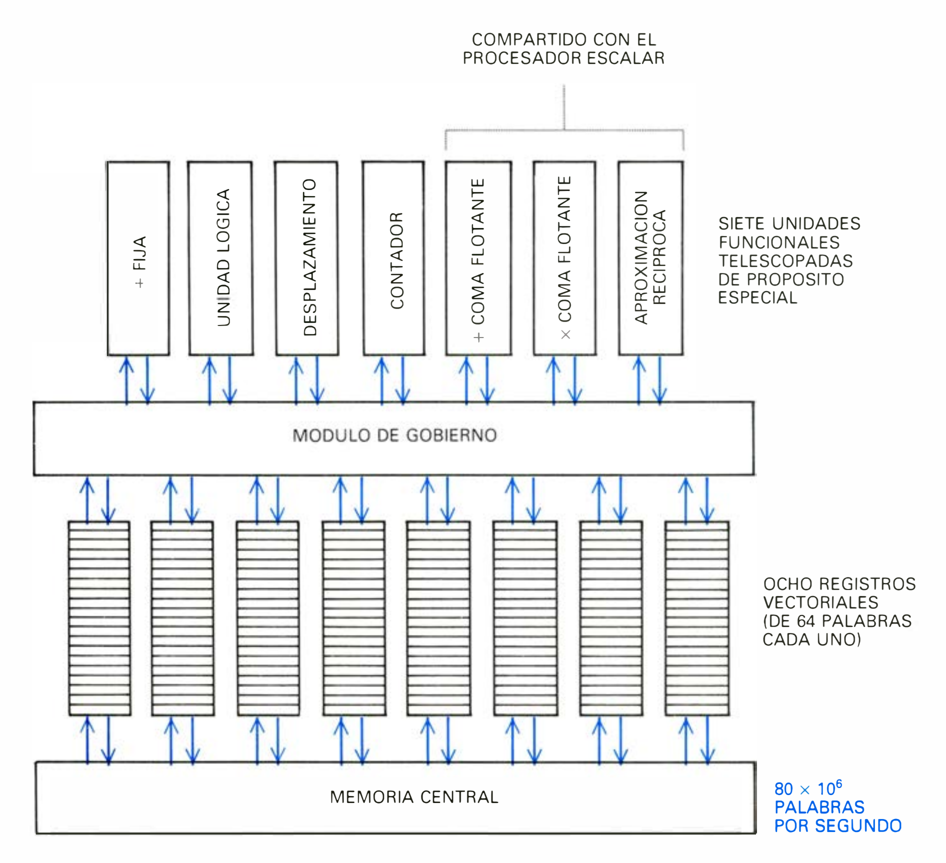
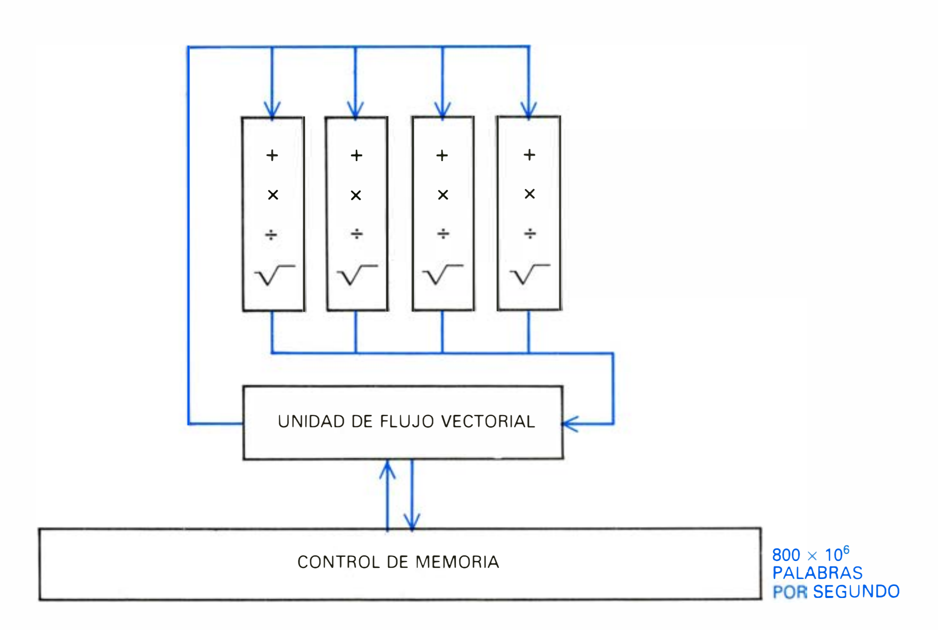


DIAGRAMA DE BLOQUES GENERICO de un moderno supercomputador vectorial, que puede interpretarse como un modelo simplificado del Cray-1 o del CYBER 205. La memoria está organizada jerárquicamente. Los dos bloques de registros corresponden a la memoria menor, seguidos en capacidad por la central, la extensión de semiconductor y la de discos. Las capacidades aproximadas aparecen en negro, mientras el ancho de banda, o velocidad de transferencia, figura en color. La expansión de memoria de semiconductor ha hecho su aparición recientemente en las instalaciones de supercomputadores al perder su puesto la tecnología de discos, incompatible con las crecientes velocidades de los procesadores (la mayor memoria de semiconductor, con una capacidad de 8 millones de palabras, se halla en un Cray-1 instalado en los Exxon Research Laboratories de Houston, Texas). En la figura, todas las unidades funcionales son telescopadas, lo que significa que las tareas se dividen en elementos ejecutables a máxima velocidad y posteriormente reunidas en un flujo continuo, razón por la cual puede realizarse una operación en coma flotante por período de reloj: 20 nanosegundos en el CYBER 205, 12,5 nanosegundos en el Cray-1. Todas las unidades funcionales pueden trabajar concurrentemente, pero no todas a máxima velocidad, puesto que comparten secciones comunes, tales como vías de datos o ciclos de acceso a memoria. Además, los saltos condicionales interrumpen el flujo uniforme de instrucciones en el correspondiente procesador. Antes de que el procesador ejecute una instrucción deberá esperar hasta estar seguro de que dispondrá de todos los recursos precisos para ello, cuando los necesite. En el Cray-1, los registros incorporan una subjerarquía propia, a la vez que el procesador vectorial contiene una memoria de registros adicional. Los procesadores vectoriales del Cray-1 y del CYBER 205 difieren en muchos otros aspectos.



PROCESADOR VECTORIAL DEL CRAY-1. De las siete unidades telescopadas de propósito especial de que consta, para la ejecución de funciones específicas tres las comparte con el procesador "escalar", que maneja operaciones no vectorizadas (el módulo contador cuenta el número de unos binarios, útil en algunos problemas). Varias de estas unidades pueden funcionar concurrentemente con distintas operaciones vectoriales. Los datos vectoriales salen de los ocho registros hacia las unidades funcionales y vuelven a los registros. El módulo de gobierno conmuta la comunicación de los operandos desde los registros hasta las unidades funcionales, y viceversa. Mientras unos registros se emplean como origen o destino de las operaciones vectoriales, otros se hallan en transferencia de datos hacia, o desde, la memoria central. En virtud de la comunicación entre registros, la circulación vectorial es corta y el tiempo de arranque breve.

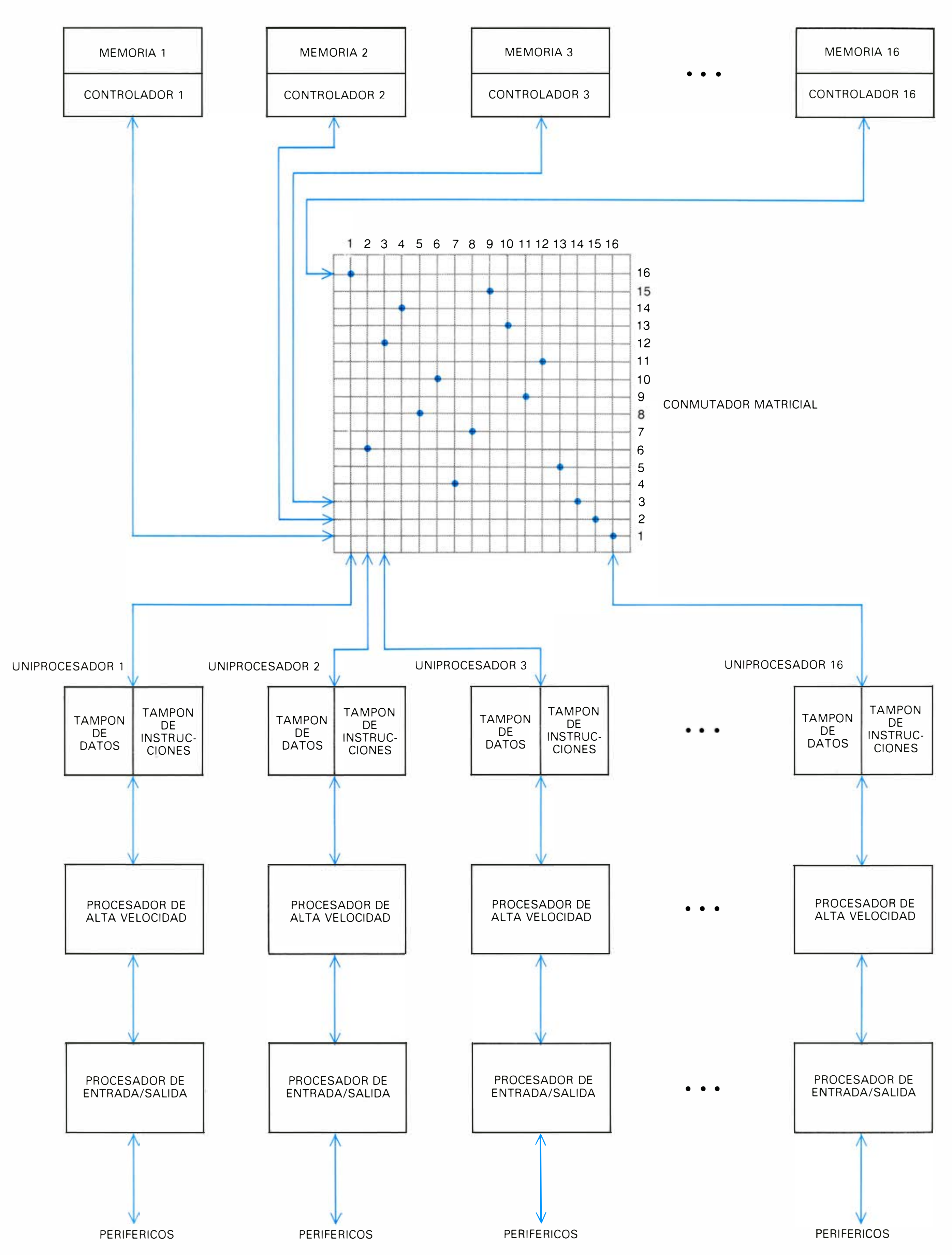


PROCESADOR VECTORIAL DEL CYBER 205. Puede incorporar hasta cuatro procesadores aritméticos telescopados de propósito general en como flotante. Cada uno de ellos puede sumar, restar, multiplicar, dividir y extraer raíces cuadradas, pero en cada instante todos se encuentran realizando la misma operación sobre diferentes pares de elementos en una única operación vectorial. La unidad de flujo vectorial maneja el tráfico entre los sistemas telescopados y la memoria central. Puesto que la memoria central se emplea tanto como origen de operandos como destino de los resultados, la longitud total de la circulación de datos es larga y el tiempo de arranque grande. Ahora bien, por ser la velocidad de transferencia a la memoria central mucho mayor que en el Cray-1 (800 millones de palabras por segundo), el CYBER 205 supera al supercomputador Cray-1 cuando la longitud del vector excede cierto valor.

extraer raíces cuadradas. Para una operación vectorial determinada, por ejemplo de suma, todos los sistemas funcionan simultáneamente, produciendo cada uno un resultado cada 20 nanosegundos. El tiempo de arranque vectorial es bastante largo: cerca de 50 ciclos, equivalentes a un microsegundo, entre el instante en que se obtiene el código de una instrucción y el momento en que aparece el primer resultado. Por tanto, el máximo rendimiento puede alcanzarse solamente con vectores largos, de forma que la pérdida debida al arranque pueda ser amortizada ejecutando numerosas operaciones en coma flotante. Para alcanzar el 50 por ciento del número máximo de megaflops, la longitud del vector debe ser, por lo menos, de 200. El CYBER 205 ofrece a los programadores una posibilidad útil de semi-precisión: cada sistema telescopado puede realizar operaciones en 32 bits al doble de velocidad que en 64.

La unidad del proceso del Cray-1 incorpora 13 unidades funcionales telescopadas de propósito especial, capaces de operar todas en paralelo. Incluye también un amplio conjunto de registros con su propia estructura jerárquica. Tres de las unidades funcionales están dedicadas a aritmética en coma flotante: una para la suma y resta, una para la multiplicación y una para el cálculo de recíprocos. Hay una diferencia importante en la manera en que el Cray-1 y el CYBER 205 manejan los vectores en el funcionamiento telescopado. En contraste con la operación directa de memoria a memoria del CYBER 205, en el Cray-1 se emplea un conjunto de registros vectoriales de ocho componentes, cada uno de los cuales es capaz de contener 64 palabras de 64 bits. Una consecuencia del trasvase registro a registro de los vectores radica en el hecho de que la gráfica eficacia/longitud del vector presenta máximos cuando la longitud es múltiplo de 64, de modo parecido a los máximos del ILLIAC IV. Sin embargo, dado que el tiempo perdido en el arranque del funcionamiento telescopado decrece rápidamente, a medida que aumenta la longitud vectorial, la eficacia del Cray-1 aumenta con dicha longitud.

Otra consecuencia de este esquema es que el sistema del Cray-1 puede resultar operativo en 50 nanosegundos (4 períodos de reloj), es decir, tan sólo una vigésima parte del que precisa el CYBER 205. Por tanto, el rendimiento del Cray-1 es superior al del CYBER 205 para vectores cortos. El cuello de botella principal en las operaciones vecto-



MULTIPROCESADOR S-1, que se encuentra actualmente en fase de construcción para la Marina de Guerra de los Estados Unidos en el Lawrence Livermore National Laboratory. Coordinará las operaciones de hasta 16 uniprocesadores, cada uno de los cuales es un computador de propósito general diseñado para ser tan potente como el Cray-1. El conmutador matricial de alta velocidad permitirá a cada procesador de S-1 acceder a cualquiera de los 16 grandes bancos de memoria. Cada uno de estos bancos está conectado a una fila de la matriz y cada uniprocesador lo está a una columna. En un instante determinado, cualquier fila puede conectarse con cualquiera de las 16 columnas. La conmutación de las conexiones se hará bajo control de programa. El primer uniprocesador, el Mark II A, está a punto de entrar en funcionamiento. El coste de éste es muy inferior al presupuestado para el Cray-1.

riales del Cray-1 está en la limitación de rapidez en las transferencias entre el procesador central y la memoria principal. Como compensación a esto, el Cray-1 dispone de una útil posibilidad: la de realizar una cadena de operaciones simultáneas sobre los elementos vectoriales mientras se hallan en los registros. La optimización del empleo del Cray-1 depende de hacer máxima dicha cadena. El CYBER 205 permite sólo una forma limitada de encadenamiento para operaciones triádicas, en las que participen dos operandos vectoriales y uno escalar.

En resumen, el Cray-1, debido a su corta duración de ciclo, es el más rápido para problemas en que dominen las operaciones escalares y con vectores cortos. El CYBER 205, merced a su gran ancho de banda en la transferencia procesador central a memoria, lo es para aquellos casos en que los vectores largos sean los preponderantes. En cuanto al ILLIAC IV, en razón del gran ancho de banda que presenta la comunicación con la memoria secundaria, era más rápido que cualquier modelo del Cray o del CYBER construido para problemas con bases de datos muy grandes.

A pesar del énfasis puesto en la utilidad de las operaciones vectoriales en la simulación de campos continuos, hay problemas importantes que no admiten una forma vectorial eficaz. Los problemas que exijan mucha búsqueda y clasificación o en los que imperen saltos condicionales no son vectorizables de forma adecuada. Sin embargo, algunos de ellos poseen otras formas de concurrencia inherente que pueden ser explotadas por los métodos de cómputo en paralelo. Los computadores vectoriales pueden ejecutar instrucciones tanto vectoriales como escalares, pero únicamente de forma secuencial única. Reciben el nombre de máquinas de instrucción única – datos múltiples, o SIMD (del inglés, "single-instruction-multiple-data"). Las computadoras secuenciales convencionales que no permiten cómputos por concurrencia se llaman también máquinas de instrucción única-datos únicos, o SISD (del inglés, "single-instruction-single-data").

Durante mucho tiempo se han estudiado las máquinas que puedan manejar múltiples instrucciones y múltiples datos –máquinas MIMD–. Se trata de un conjunto de procesadores de propósito general, capaces, cada uno de ellos, de ejecutar una serie de instrucciones independientemente de los demás pero disponiendo a la vez de un cierto nivel de comunicación entre ellos, de suerte que puedan colaborar en la resolución de un determinado problema. En varias universidades se han construido pequeños MIMD experimentales. Los primeros supercomputadores MIMD funcionarán a pleno rendimiento este mismo año.

Una forma obvia de construir un MIMD consiste en reunir varios SIMD y una gran memoria rápida mediante una red de conmutación de alta velocidad. Otra posibilidad es dotarlo de un uniprocesador de arquitectura telescopada utilizando la multiplexación, o tiempo compartido, del soporte físico ("hardware") entre las series de instrucciones.

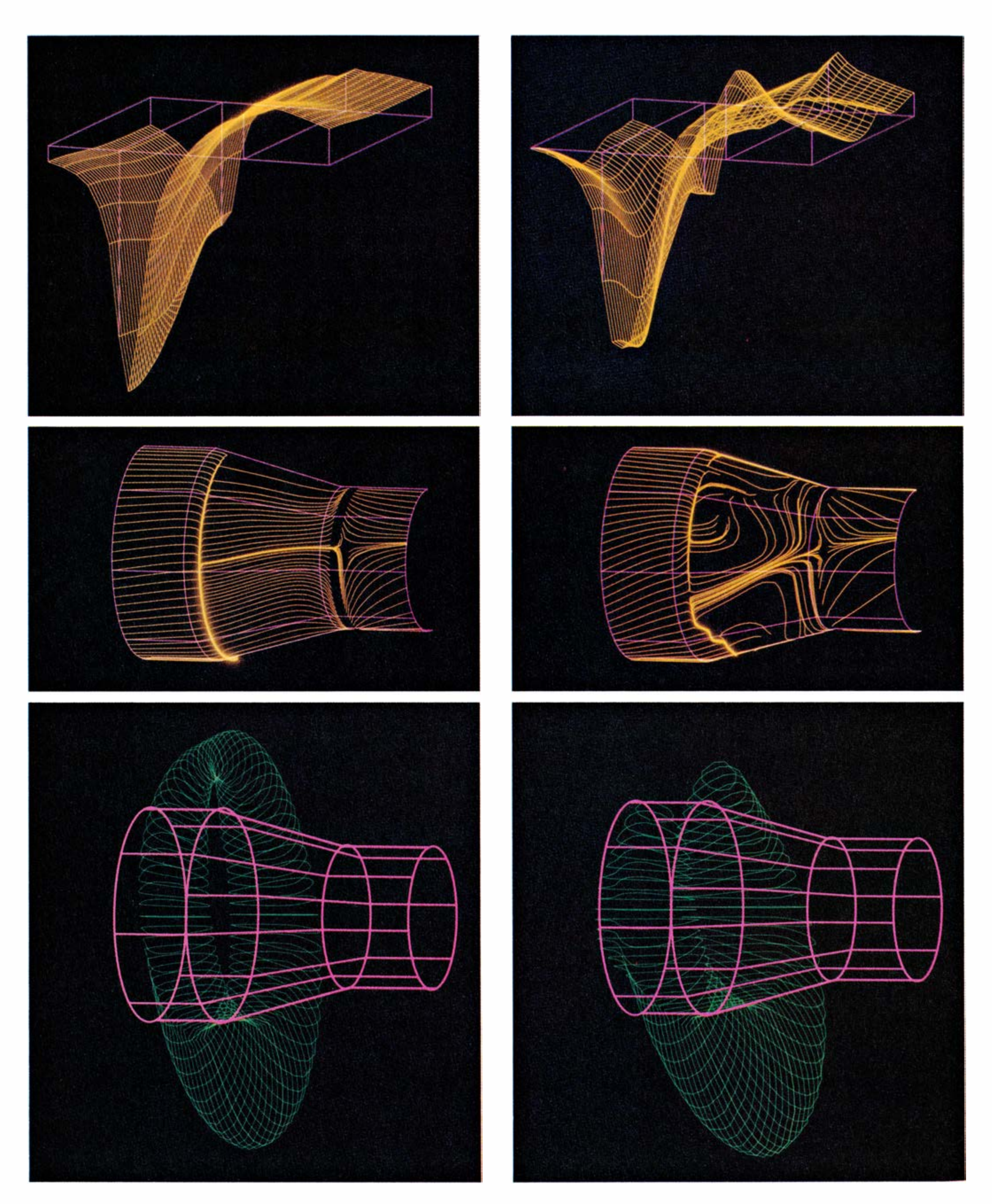
Un ejemplo interesante de este tipo lo constituye el HEP (heterogeneous element processor), un sistema de multiproceso desarrollado por la firma Denelcor, Inc., de Denver, Colorado, para los Laboratorios de Investigación de Balística del Ejército. El uniprocesador básico del sistema, el módulo del elemento de proceso (PEM), puede manejar múltiples procesos mediante reparto de tiempo del soporte físico asignado al control y ejecución de las instrucciones. Periódicamente se ejecuta una instrucción de cada proceso; la lógica está dispuesta de forma telescopada, con lo que en cada instante los distintos procesos se hallan en distintas fases de su resolución. En cada período de reloj puede iniciarse la ejecución de una nueva instrucción, completándose, en la mayoría de los casos, en un total de 8 períodos. A pesar de que cada proceso se atiende, como máximo, una vez cada 8 ciclos, los distintos procesos pueden constituir otras tantas subtareas de un único programa, de forma que la velocidad máxima de ejecución es realmente de una instrucción por ciclo.

Para la sincronización entre los procesos cooperativos se emplean "semáforos" de soporte físico ligados a las zonas de memoria que indican cuándo pueden almacenarse allí datos suministrados por un proceso o si ya contienen datos en espera de ser empleados. En el PEM pueden hallarse en realización hasta 128 procesos; un mínimo de 8 de ellos se emplean para utilizar al máximo el soporte físico y la velocidad de ejecución de las instrucciones. Los circuitos del HEP están fabricados, exclusivamente, con integración en media escala, con lo que el período del reloj es de unos 100 nanosegundos, lo que supone una velocidad máxima de 10 millones de instrucciones por segundo. Dado que la máquina no está preparada para manejar instrucciones vectoriales, el número de megaflops es algo inferior a 10.

El mayor sistema HEP podría llegar a conectar 16 PEM con una memoria de 128 millones de palabras de 64 bits mediante un conmutador de datos de diseño muy avanzado. Los procesos correspondientes a un único programa pueden distribuirse dinámicamente entre varios HEP, puesto que éstos poseen una interesante propiedad de seguridad: cuando uno o más PEM están fuera de servicio para reparaciones o mantenimiento, el resto del sistema puede continuar trabajando, aunque con una velocidad proporcionalmente menor. Un HEP con 16 PEM puede llegar a ejecutar 160 millones de instrucciones por segundo, es decir, el doble de la máxima velocidad del Cray-1. Sin embargo, dado que su precio es mucho mayor, no resulta competitivo para problemas que puedan ser vectorizados, los cuales resultan más adecuados para la arquitectura del Cray-1. En principio, sin embargo, no parece haber ningún obstáculo para dotar la arquitectura del HEP con tecnologías integradas más rápidas y densas, con las que poder multiplicar por ocho la velocidad y aumentar considerablemente la razón precio/eficacia.

Los nuevos supercomputadores en curso de construcción o en etapa de diseño muestran una tendencia a aumentar el multiproceso. Mientras que existen límites fundamentales al punto en que un proceso puede llegar a segmentarse para pemitir concurrencias telescopadas, la multiplicación del poder de cálculo mediante adición de procesadores paralelos está limitada únicamente por su costo. La mayoría de los diseñadores están de acuerdo en admitir que todo multiprocesador debería incorporar un pequeño número de uniprocesadores, cada uno muy potente de por sí. Aunque el precio de un sistema de multiproceso aumenta linealmente con el número de uniprocesadores, la aceleración del rendimiento es menor debido al tope de comunicaciones y a las inevitables limitaciones de la utilización de los uniprocesadores.

El éxito en la carrera por el desarrollo de los supercomputadores dependerá fuertemente de la reducción del tiempo preciso para diseñar y construir una máquina capaz de incorporar los avances tecnológicos. Es bien sabido que los ordenadores pueden emplearse en diseñar mejores ordenadores. Como un primer paso para el diseño de un supercomputador de multiproceso avanzado, con destino a la Marina estadounidense, el grupo del Lawrence Livermore National Laboratory, que trabaja en el proyecto S-1, desarrolló el sistema apodado SCALD ("*s*tructured *c*omputer-*a*ided *l*ogic *d*esign", diseño lógico es-



DIVERSAS REPRESENTACIONES VISUALES pueden obtenerse a partir de la solución de un problema complejo por computador. En este caso, el problema es el mismo que el mostrado en la figura de la página 71: el modelo de presión de aire que circula alrededor de la parte posterior de un cohete que avanza a una velocidad por debajo justo de la del sonido. El ILLIAC IV calculó el flujo para diversos ángulos de ataque, esto es, para diversos ángulos de inclinación entre el eje del cohete y su dirección de vuelo. La solución para cada ángulo de ataque exigió alrededor de 18 horas de computación. Las figuras de las dos columnas, programadas por Adrian Borja, ilustran tres formas de representar una pequeña parte de los resultados para dos ángulos de ataque: 4 grados (*izquierda*) y 12 grados (*derecha*). Las gráficas en forma de velo de la fila superior muestran los cambios de la presión del aire sobre la mitad superior de la parte posterior del cohete. Las figuras rayadas de la fila central muestran los cortes superficiales: el flujo de aire adyacente a la superficie posterior del cuerpo del cohete. Con un ángulo de ataque de 12 grados la dirección del flujo se invierte. Las muestras de la fila inferior delinean los contornos en forma de herradura de la superficie sónica, en la que el flujo de aire alcanza la velocidad del sonido. Las tres figuras de la columna de la derecha, correspondiente al ataque de 12 grados, siguen el mismo patrón que el mostrado como una superficie continua transparente en la página 71.

tructurado asistido por ordenador). El SCALD consta de unas 60.000 líneas de código de lenguaje PASCAL de alto nivel. El sistema permite al ingeniero expresar sus conceptos de diseño, en cualquier aspecto del diseño lógico, con el mismo lenguaje gráfico abstracto que emplea en sus razonamientos y en su relación con otros ingenieros. De la parte más tediosa del diseño y sus pruebas se encarga la máquina.

Con el sistema SCALD, un pequeño grupo de ingenieros de Livermore realizó, en menos de un año, un detallado diseño de un potente calculador científico tipo SIMD, llamado Mark IIA. Concebido como un uniprocesador apto para un eventual multiprocesador, el Mark IIA posee una arquitectura rica y flexible. El grupo de Livermore pretende construir una máquina que pueda compararse favorablemente con el Cray-1. Resulta realmente difícil, sin embargo, predecir los rendimientos relativos de ambas máquinas, debido a que sus jerarquías de memoria están estructuradas muy diferentemente, en términos de su capacidad y ancho de banda en cada nivel. El coste del Mark IIA resultará sensiblemente menor que el del Cray-1, merced a sus nuevos circuitos integrados de gran escala. Así, la memoria principal estará construida con tecnología MOS con 65.536 bits por pastilla, en comparación con los 4096 bits del Cray-1. Un multiprocesador S-1 podría contener hasta 16 uniprocesadores Mark IIA y cientos de millones de palabras de memoria compartida.

Para multiprocesadores tipo MIMD, el S-1 y el HEP por ejemplo, la máxima velocidad alcanzable, a partir de sus procesadores integrantes, depende, en definitiva, de la naturaleza del problema que deba resolverse. Para el caso de problemas relativos a campos continuos, altamente iterativos, la explotación eficiente de tales máquinas exige el desarrollo de nuevos algoritmos. Por otra parte, algunas aplicaciones, tales como el procesamiento de imágenes en tiempo real, admiten de forma natural el multiproceso, puesto que pueden ser telescopados sin dificultad: mientras el procesador n.º 1 realiza el primer paso de reducción en el bloque n de datos, el n.º 2 efectúa el segundo paso de cómputo en el bloque $n-1$, el n.º 3 se halla en la tercera etapa del $n-2$, y así sucesivamente.

Las comunicaciones entre un ordenador y el medio externo, conocidas como operaciones de entrada/salida (E/S), son muy lentas en comparación con las velocidades de procesamiento de la máquina, dado que los dispositivos periféricos poseen componentes mecánicos y que el tiempo de respuesta humano es elevado. Para liberar a los supercomputadores del control del elevado tráfico de entrada/salida, y emplearlos únicamente en mantener la velocidad de cálculo, la mayor parte de ellos están equipados con procesadores de E/S y con uno o más ordenadores satélite llamados terminales (front-end). Los últimos modelos del Cray-1 disponen de una serie de procesadores E/S, que no son sino computadores rápidos de pequeña longitud de palabra y con un limitado conjunto de instrucciones, responsables del manejo del tráfico de datos a alta velocidad entre la memoria central y la secundaria, otros periféricos y los ordenadores terminales. Estos actúan como secretarias del poderoso supercomputador. Los ordenadores terminales pueden, a su vez, ser auténticos computadores científicos, capaces de llevar a cabo parte de la carga del proceso de datos, en particular reducir una gran masa de resultados a una forma humanamente inteligible. Por ejemplo, la solución completa de una simulación de flujo aerodinámico puede contener un millón de valores numéricos, los cuales sobrepasarían la capacidad del ingeniero de no ser presentados en forma gráfica, quizá como una película.

Algunos centros de cálculo han desarrollado sofisticadas redes de datos para intercomunicar varios supercomputadores, ordenadores terminales, computadores de propósito general, grandes bancos de datos y equipos de visualización, de forma que todo usuario puede acceder rápidamente a los elementos más idóneos para su problema. Aunque el número de supercomputadores es aún reducido, cada vez son más accesibles gracias a las redes nacionales e internacionales que comunican a las instituciones científicas.

Los usuarios de los superordenadores exigen potencias de cálculo cada vez mayores, para poder abordar los problemas de cómputo que se presentan en las fronteras del conocimiento de todas las disciplinas. Los ingenieros de ordenadores tienen distintas ideas para satisfacer tales demandas. Es muy probable que en los próximos años veamos el acelerado desarrollo y evaluación de nuevas soluciones, espoleadas por los rápidos cambios de la tecnología, las avanzadas técnicas del diseño asistido por ordenador y por una amplia disponibilidad de las supermáquinas sometidas a prueba por una crítica comunidad de usuarios.

Formación de un sistema nervioso simple

Rastreando la genealogía de células nerviosas durante el desarrollo de sanguijuelas, enanas y gigantes, se han obtenido las primeras claves del funcionamiento del sistema nervioso en el adulto

Gunther S. Stent y David A. Weisblat

El sistema nervioso nos plantea dos de los interrogantes más fascinantes de la biología contemporánea. ¿Cómo actúan las redes de neuronas que generan la conducta animal? ¿De qué forma establecen esas células nerviosas sus conexiones específicas durante el desarrollo embrionario del animal a partir del óvulo fecundado? La segunda cuestión no puede abordarse sin considerar antes la primera: la anatomía y fisiología del sistema nervioso del adulto representan el punto final del desarrollo del sistema nervioso. Un conocimiento anatómico y funcional detallado constituye un requisito imprescindible, aunque no sea más que para plantear las cuestiones correctamente. Conocimiento que nos lo puede ofrecer el examen de animales que poseen un sistema nervioso simple. La sanguijuela, por ejemplo, un anélido emparentado con la lombriz de tierra, que se alimenta de sangre.

La sola mención de su nombre provoca en muchos la repugnancia que mostrara Charlie Allnutt (personificado por Humphrey Bogart) en *La reina de Africa*. Saliendo de una ciénaga, Allnutt descubre los parásitos adheridos a su piel, y exclama: "¡*Abomino* las sanguijuelas!". Pero esos gusanos presentan ciertos aspectos positivos que, sin duda, Allnutt desconocía. Mientras succionaban su sangre, las sanguijuelas inyectaban en la corriente sanguínea sustancias anticoagulantes capaces de reducir el riesgo de trastornos cardiovasculares.

Por otra parte, centrándonos en nuestro tema, la organización extraordinariamente simple del cuerpo de la sanguijuela convierte a este animal en un modelo atractivo para el biólogo experimental. Su cuerpo cilíndrico está formado por 32 segmentos similares, con la peculiaridad de que, estudiando uno solo de éstos, podemos interpretar la organización del animal entero. De los 32 segmentos, los cuatro proximales constituyen las estructuras especializadas de la cabeza, entre ellas un par de ojos en el dorso o superficie superior y una ventosa proximal en la parte ventral, o superficie inferior. Los siete segmentos caudales constituyen estructuras especializadas, tales como el ano y la gran ventosa distal. La anatomía de los 21 segmentos intermedios es muy estereotipada. Cada segmento contiene una serie de órganos viscerales; vasos circulatorios, riñones e intestino. La piel de cada segmento está subdividida siempre en un mismo número de anillos, de los cuales el medial posee una serie de órganos sensitivos distribuidos a lo largo de la circunferencia del cuerpo. La pared corporal de cada segmento está rodeada por músculos circulares que pueden constreñir la masa corporal. Y si descendemos en este corte transversal hipotético, veremos los músculos longitudinales, cuya contracción acorta el tubo corporal.

El sistema nervioso de la sanguijuela refleja la organización segmentaria de su cuerpo. Consta de 32 ganglios interconectados que forman el cordón nervioso ventral. En cada ganglio hay 200 pares de neuronas con simetría bilateral y unas cuantas desapareadas. En cada segmento, el ganglio está unido a la pared corporal y a los órganos internos por dos pares bilaterales de nervios y, a los ganglios vecinos, por haces de nervios de conexión. Estudios exhaustivos del sistema nervioso de la sanguijuela, llevados a cabo por John G. Nicholls y sus discípulos, en la Universidad de Stanford, han puesto de manifiesto que la anatomía de los ganglios segmentarios es lo bastante estereotipada de uno a otro segmento y suficientemente invariable de un animal a otro para que muchas de las neuronas puedan identificarse sin mayor dificultad.

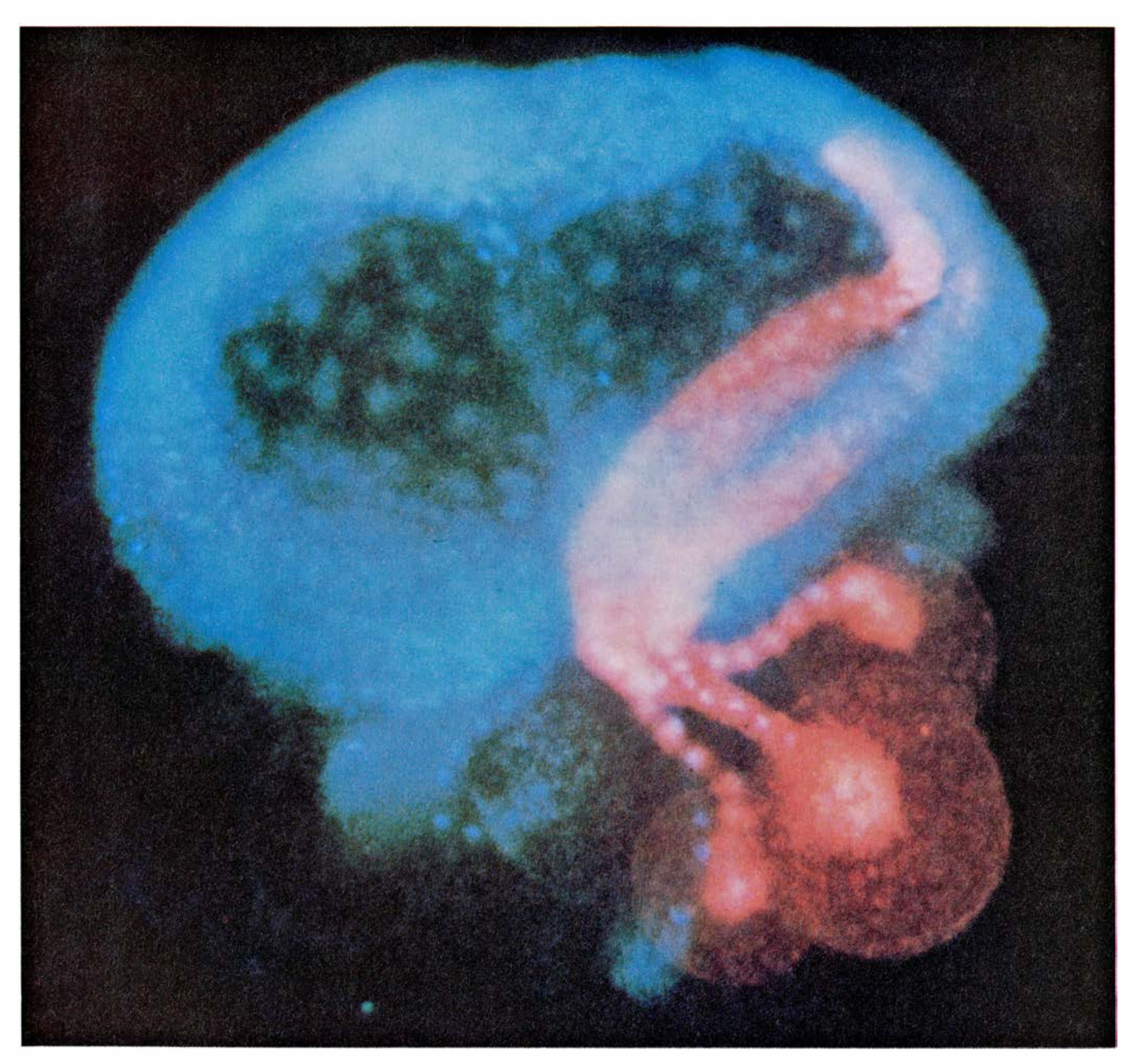
Las neuronas de los ganglios segmentarios admiten la introducción de microelectrodos para registrar la actividad eléctrica. También puede inyectarse un colorante o una sustancia fluorescente por medio de una micropipeta para revelar la anatomía detallada de una neurona individual. Estas técnicas han permitido establecer el patrón de conexiones interneurales y explicar algunos actos simples de la conducta refleja, así el encogimiento del cuerpo en respuesta a un estímulo táctil, e incluso algunos movimientos más complejos, como la actividad cardíaca.

La conducta de mayor complejidad descrita hasta la fecha, a través de la identificación de las neuronas implicadas y sus conexiones, es la natación. Los experimentos llevados a cabo en nuestro laboratorio de la Universidad de California en Berkeley, de 1971 a 1977, por William B. Kristan, Jr., Carol Ort, Otto Friesen, Margaret Poon y Ronald Calabrese, mostraron que el ritmo contráctil de los músculos longitudinales responsable de los movimientos de natación se generaba en una serie de 12 pares bilaterales de neuronas motoras de cada ganglio segmentario que exhibían una actividad rítmica. El ritmo de las neuronas motoras dependía de cuatro pares bilaterales de interneuronas, que integran el oscilador natatorio central.

Se conocen, con bastante profundidad, los elementos funcionales del sistema nervioso de la sanguijuela, hasta el punto de que el modelo nos da pie a plantearnos cuestiones específicas sobre los procesos del desarrollo. ¿Cómo se establecen los circuitos nerviosos responsables del control de la natación? ¿Empiezan a formarse las cone-

xiones desde fuera, de modo que en primer lugar se conectan los músculos longitudinales a las neuronas motoras, luego éstas a las interneuronas y, finalmente, éstas al oscilador central? ¿Tiene lugar, por contra, la formación del circuito en sentido inverso, de dentro afuera? ¿Acaso se establecen los tres niveles de conexión al mismo tiempo? Puede que el circuito no se desarrolle por establecimiento de conexiones específicas, sino que se forme por selección, en una red neuronal con múltiples conexiones, de aquellas que resultan funcionalmente adecuadas.

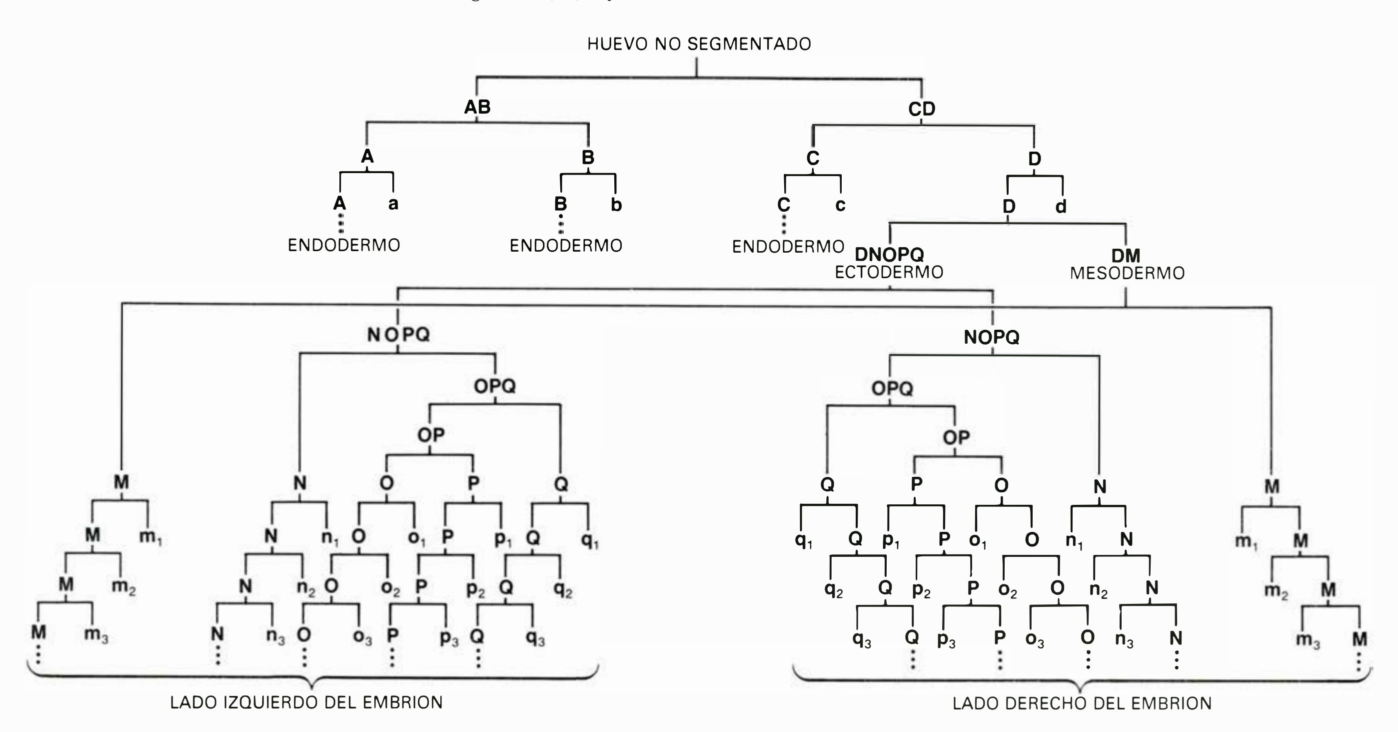
Hace siete años, con la esperanza de responder alguna de estas cuestiones, iniciamos un proyecto de investigación a largo plazo sobre el desarrollo del sistema nervioso de la sanguijuela. Nos fijamos la primera meta: hallar una especie adecuada para su cría en el laboratorio que proporcionara un suministro permanente de embriones. A tal fin, Roy T. Sawyer se encargó de examinar muchas especies hasta dar con dos candidatas para nuestro proyecto: *Helobdella triserialis* y *Haementeria ghilianii*. La primera, nativa de California, se alimenta de los líquidos corporales de los caracoles; alcanza, en su forma adulta, los dos centímetros de longitud y su ciclo de reproducción (de huevo a huevo) dura ocho semanas. *Haementeria*, que procede de la Guayana francesa, se alimenta de sangre de mamíferos; alcanza una longitud de 50 centímetros y su ciclo es de 10 meses.



ESTADIO TEMPRANO del desarrollo embrionario de una sanguijuela enana, *Helobdella triserialis*. Se ha utilizado aquí una técnica de marcaje ideada por los autores y sus colegas, de la Universidad de California en Berkeley. En una fase inicial del desarrollo del embrión se inyectó en una sola célula, mediante una micropipeta, un colorante rojo fluorescente acoplado a un polipéptido sintético. Se fijó entonces el embrión, se le eliminó el vitelo nutritivo y se trató con una sustancia fluorescente azul, específica para el ADN. Luego se fotografió la preparación a través de un microscopio. Los puntos azules corresponden a núcleos de células que han empezado a unirse para formar las estructuras embrionarias en forma de media luna, denominadas bandas germinales. Dichas células se concentran particularmente en la parte anterior del embrión, correspondiente a la cabeza (*arriba*). Las tres células de gran tamaño situadas en la parte inferior derecha descienden directamente de la célula a la que se inyectó inicialmente la sustancia fluorescente roja. De cada una de estas células se origina una cintilla de células hijas, denominadas células progenitoras, que se proyectan en la banda germinal derecha. La totalidad del sistema nervioso de la sanguijuela adulta deriva de las cintillas de células progenitoras. La ampliación es, aproximadamente, de unos 300 diámetros.

ESTADIO 1 — TELOPLASMA
ESTADIO 2 — AB, CD, MEMBRANA DEL HUEVO, TELOPLASMA
ESTADIO 3 — A, B, C, D
ESTADIO 4 — MICROMEROS, A, B, C, DM, DNOPQ
ESTADIO 5 — A, B, C, M, NOPQ, NOPQ, M
ESTADIO 6 — A, B, C, M, N, O, P, Q, N, O, P, Q, M
ESTADIO 7 (INICIAL) — M, N, O, P, Q, N, O, P, Q, M, CINTILLA DE CELULAS PROGENITORAS
ESTADIO 7 (MEDIO) — BANDAS GERMINALES
ESTADIO 8 (INICIAL) — CASQUETE MICROMERICO
ESTADIO 8 (MEDIO) — PLACA GERMINAL, CORDON NERVIOSO
ESTADIO 8 (FINAL) — CORDON NERVIOSO, BORDE LATERAL DE LA PLACA GERMINAL
ESTADIO 9 — MEMBRANA DEL HUEVO
ESTADIO 10 — GANGLIOS CAUDALES FUSIONADOS, ANILLOS, GANGLIOS CEFALICOS FUSIONADOS

PRIMEROS DIEZ ESTADIOS del desarrollo embrionario de *Helobdella*, representados esquemáticamente en esta ilustración, desde la etapa de huevo (*estadio 1*) hasta el tubo corporal prácticamente cerrado ya (*estadio 10*). Los dos estadios fundamentales (*7* y *8*) se han subdividido en varias fases. Los esquemas de las dos hileras de arriba muestran la futura superficie dorsal, o superior, del embrión. Los esquemas de la hilera inferior muestran la superficie lateral. Al comienzo del estadio 4, las células designadas *A, B, C* y *D* dan lugar a cuatro células de pequeño tamaño denominadas micrómeros, situadas en el polo dorsal. La proliferación subsiguiente de tales células produce la expansión del casquete microméricol (*superficie azul*) que llega a envolver al embrión primitivo. En los estadios 9 y 10, las células de la placa germinal (*superficie roja*) proliferan extendiéndose hacia la región dorsal del embrión. Finalmente, los bordes laterales izquierdo y derecho de la placa germinal establecen contacto en la línea media dorsal, cerrándose el cilindro corporal.

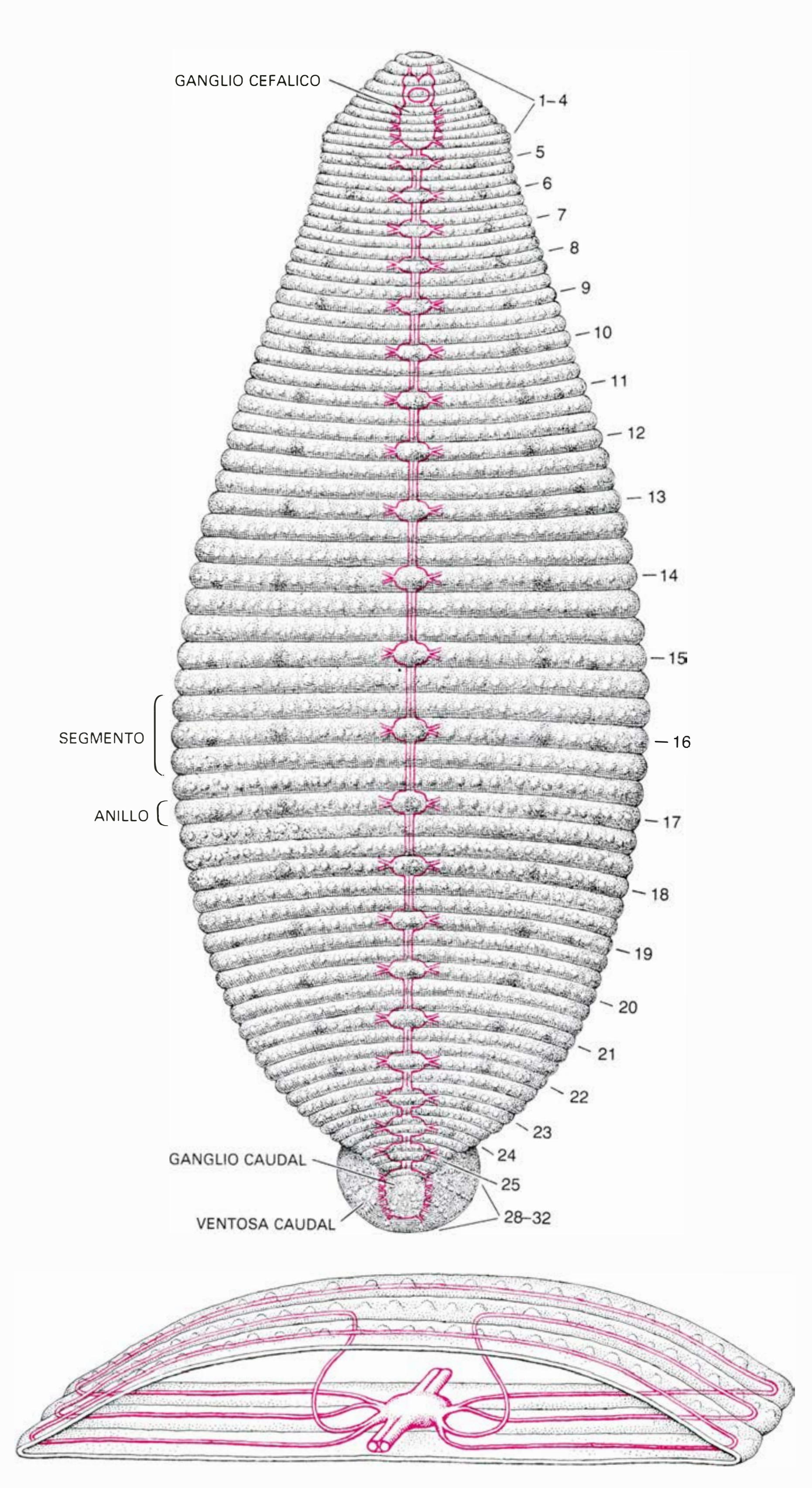


GENEALOGIA DE LAS CELULAS del embrión de *Helobdella*, en una representación de "árbol familiar". Las células del linaje se han designado con letras minúsculas y se corresponden con las letras mayúsculas de los teloblastos de los que derivan. Se ha indicado también la separación del embrión en tres capas de tejidos germinales. Las células descendientes de las células *A, B* y *C* dan lugar al endodermo (la capa más interna), las células descendientes de *DNOPQ* al ectodermo (la capa más externa) y las descendientes de *DM*, al mesodermo (la capa intermedia). Se han omitido las divisiones celulares que originan el casquete micromérico. La simetría bilateral del embrión se establece cuando las células progenitoras *DM* y *DNOPQ* experimentan división.

La rápida propagación, fácil cría y vitalidad del embrión de *Helobdella* la hacen muy adecuada para estudios embriológicos. Su breve tamaño constituye, empero, un inconveniente para la investigación neurofisiológica. Dificultad que salva el enorme tamaño de *Haementeria*, cuyo sistema nervioso y neuronas precursoras se muestran accesibles para obtener registros eléctricos intracelulares y aplicar otras técnicas que requieran la penetración de células individuales. Sin embargo, en su contra cuentan el largo tiempo de generación, las condiciones de cultivo, más difíciles, y la fragilidad de sus embriones. Se ha procurado sacar partido, obviamente, de las distintas ventajas que presentan ambas especies, que, además, pertenecen a la misma familia. Significa ello que, a pesar de su disparidad en tamaño y forma de vida, son semejantes en su organización corporal adulta y en su desarrollo embrionario. En muchos aspectos, las dos pueden considerarse equivalentes desde el punto de vista embriológico y ambas proporcionan un horizonte más amplio a la investigación que cada una de ellas por separado.

Helobdella y *Haementeria* ponen huevos ricos en vitelo nutritivo, cuyos diámetros miden 0,5 y 2,5 milímetros, respectivamente. Los huevos se depositan en grupos; cada grupo queda incluido en un capullo transparente, que permanece unido a la pared ventral de la sanguijuela. El desarrollo embrionario comienza inmediatamente después de la puesta y dura dos semanas, en *Helobdella*, y un mes, en *Haementeria*. El vitelo nutritivo proporciona las sustancias necesarias para el desarrollo embrionario. Tan pronto como se consume el vitelo, el individuo joven, que es similar al adulto en su forma, aunque de menor tamaño, ingiere su primer alimento del animal que parasita. En el curso del crecimiento y maduración postembrionarios, las células de la sanguijuela aumentan en tamaño y, en menor medida, en número.

Ya en el siglo XIX los pioneros de la moderna embriología experimental centraron su atención en estos anélidos. Hacia 1880, Charles O. Whitman, quien fuera el primer director de los laboratorios de biología marina de Woods Hole, Massachusetts, describió la sucesión exacta de divisiones celulares a través de las cuales se formaba el embrión de la sanguijuela a partir de un huevo fecundado. Con ello realizó el primer análisis del desarrollo de un linaje celular. Partiendo de sus observaciones, Whitman propuso que cada cé-



ORGANIZACION SEGMENTARIA del cuerpo de la sanguijuela, evidente en el dibujo de la parte superior, que representa una visión dorsal de *Haementeria ghilianii*, una especie gigante nativa de la Guayana francesa. En contraste con *Helobdella*, que a duras penas mide los dos centímetros en su forma adulta, *Haementeria* alcanza hasta 50 centímetros. No obstante, ambas especies son muy parecidas en múltiples aspectos y pueden considerarse equivalentes para determinadas investigaciones. La piel dorsal de los segmentos medios en ambas especies está subdividida en tres anillos. En la parte inferior de la ilustración se ha representado, en sección, uno de los segmentos. En color aparecen los principales componentes del sistema nervioso de la sanguijuela. El cordón nervioso ventral consta de 32 ganglios segmentarios. Los siete ganglios distales se fusionan para formar el ganglio caudal, de gran tamaño. Los cuatro proximales están fusionados y forman la parte inferior del ganglio cefálico. (La parte superior de este ganglio tiene un origen diferente.) Los nervios segmentarios se muestran con mayor detalle en el segmento seccionado de la parte inferior de la figura, que se ha dibujado en perspectiva. (La ilustración es de Tom Prentiss.)

lula embrionaria, más el conjunto de células descendientes de la misma, desempeñaba un papel específico, predestinado, en el desarrollo ulterior. La hipótesis, revolucionaria en aquel momento, constituye aún un tema central en las discusiones sobre desarrollo embrionario.

El interés de los embriólogos por esta especie declinaría luego, a principios de nuestro siglo, después de que Whitman dirigiera su atención hacia la conducta animal y fundara las bases de la etología. Por aquellas fechas, su discípulo Thomas Hunt Morgan iniciaba las investigaciones sobre la genética de la mosca de la fruta *Drosophila melanogaster*, revolucionando la ciencia de la herencia. El reciente resurgimiento de los estudios sobre la embriología de la sanguijuela se debe a Sawyer y a Juan Fernández, de la Universidad de Chile.

Cuando el huevo fecundado se acerca a la primera división celular, se aprecia en cada polo una región del citoplasma menos coloreada, denominada teloplasma, que se distingue del resto del vitelo coloreado que llena la mayor parte del huevo. Uno de los polos corresponde a la futura superficie dorsal del embrión y, el otro, a la superficie ventral. La primera división origina dos células de gran tamaño, las llamadas *AB* y *CD*; la mayor parte del teloplasma se sitúa en la célula *CD*. La segunda división produce cuatro células: *A*, *B*, *C* y *D*; esta última célula recibe el teloplasma entero. Cuando se divide la *D*, una de sus células hijas, *DNOPQ*, se sitúa más dorsalmente y recibe el teloplasma del polo dorsal; la otra célula, denominada *DM*, se encuentra más ventralmente y recibe el teloplasma del polo ventral.

En este estadio pueden distinguirse en el embrión tres capas de células de tejido germinal; las células descendientes de *A*, *B* y C darán lugar al endodermo (la capa más interna), las descendientes de *DNOPQ* al ectodermo (la más externa) y las descendientes de *DM* al mesodermo (la capa media). Las dos divisiones subsiguientes establecen la simetría bilateral del embrión: la célula *DM* se divide y origina las células *M* izquierda y derecha, y las *DNOPQ* se dividen para dar lugar al par de células *NOPQ*, situadas a cada lado de la futura línea media. Tres divisiones ulteriores de los pares de células *NOPQ* producen cuatro pares de células bilaterales, llamadas *N*, *O*, *P* y *Q*.

Los pares de células *M*, *N*, *O*, *P* y *Q*, entre los que se ha dividido el teloplasma del huevo, se denominan teloblastos. En cuanto se forman, los teloblastos experimentan una serie de divisiones muy desiguales. A través de estas divisiones, cada teloblasto origina una cintilla de pequeñas células progenitoras en las que se encuentra la mayor parte del teloplasma, pero muy poco vitelo del teloblasto. Las cintillas de células progenitoras producidas por los cinco teloblastos a cada lado de la línea media crecen y se funden, originando un par de crestas prominentes, las bandas germinales izquierda y derecha. La cintilla de células germinales descendiente de las células *M* se encuentra situada por debajo de las otras cuatro cintillas.

A medida que se producen más células progenitoras, las bandas germinales izquierda y derecha avanzan hacia la región frontal del embrión, a lo largo de la futura superficie dorsal, convergiendo en el lugar donde se formará la cabeza. Al mismo tiempo, la parte media de las bandas germinales que experimentan alargamiento se prolonga circunferencialmente hacia la futura superficie ventral. Las bandas germinales izquierda y derecha confluyen, por último, en la línea media ventral. La coalescencia empieza a nivel de la futura cabeza y continúa, como una cremallera, hacia la cola.

Al completarse la coalescencia, las bandas germinales han dado lugar a la estructura denominada placa germinal, situada en la línea media ventral. En el curso del desarrollo subsiguiente, las células de la placa germinal se dividen para alumbrar las células precursoras de los tejidos del organismo adulto. La proliferación celular resulta en un engrosamiento gradual y una expansión circunferencial de la placa germinal, sobre la superficie del embrión y hacia la región dorsal. Finalmente, los bordes izquierdo y derecho de la placa germinal se reúnen en la futura línea media dorsal, cerrando el cilindro que constituye el cuerpo de la sanguijuela.

Al comienzo de este proceso de expansión, la placa germinal se divide en una serie de bloques tisulares, separados por tabiques. Cada bloque corresponde a lo que luego será un segmento corporal. La segmentación comienza en la región frontal y progresa distalmente hasta completarse con la formación del segmento 32, momento en que la placa germinal en expansión cubre un tercio de la superficie ventral.

Al mismo tiempo, se desarrolla el intestino embrionario, que deriva de las células precursoras *A*, *B* y *C* del endodermo. El intestino aparece al principio como un cilindro alargado lleno de vitelo de las células *A*, *B* y *C*. Ulteriormente, el cilindro queda dividido en segmentos por una serie de constricciones que se corresponden con los tabiques y que tal vez sean consecuencia de su crecimiento. Como resultado de este pro-



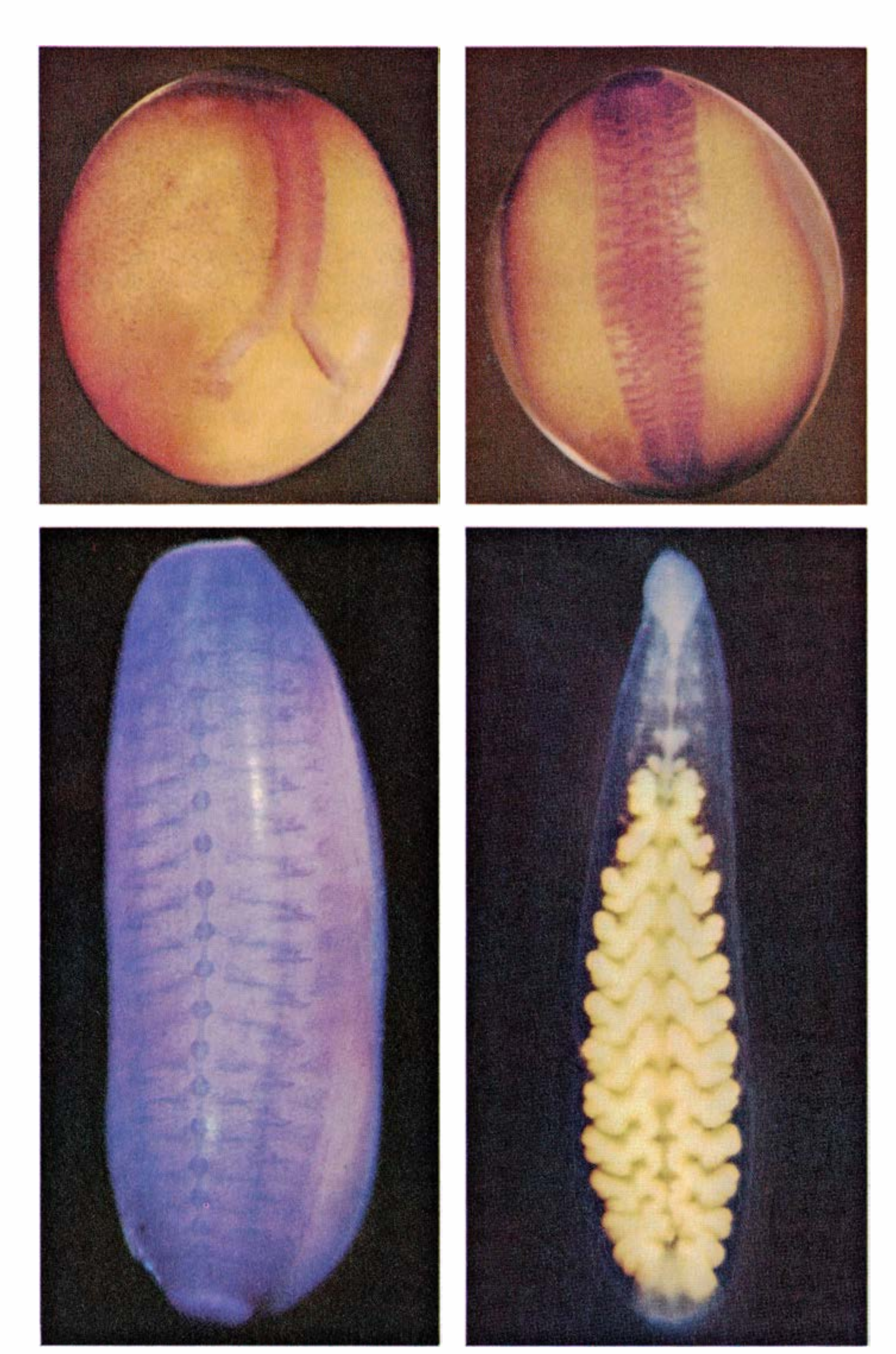
HUEVOS de la sanguijuela enana *Helobdella*, ampliados unos 50 diámetros en esta microfotografía. Algunos huevos no han empezado todavía a dividirse; otros se encuentran en su primera división y los hay, por fin, que la acaban de completar. En dos de los huevos que no se han dividido se aprecia el teloplasma como una mancha de color azul pálido en el polo del huevo. En dos de los huevos que han experimentado división, el teloplasma ha pasado a la célula hija de mayor tamaño, denominada *CD*.

ceso se desarrollan en cada segmento abdominal un par de expansiones del intestino, denominadas ciegos. La segmentación del intestino concluye al mismo tiempo que se cierra el tubo corporal; el embrión se nos ha convertido ya en una sanguijuela joven.

La segmentación progresiva de la placa germinal y el desarrollo secuencial antero-posterior se manifiestan también en la formación del sistema nervioso. La primera indicación de la presencia de tejido nervioso es la aparición de pares de masas celulares hemisféricas, a cada lado de la línea media de la placa germinal. Cada par de masas celulares constituye el "primordio" de un ganglio segmentario. Secuencialmente, experimentan coalescencia en la línea media ventral para formar ganglios globulares que contienen, aproximadamente, el mismo número de neuronas que el adulto. Cuando la placa germinal en expansión cubre alrededor de la mitad de la superficie ventral, se han formado ya los 32 ganglios. En un comienzo, cada ganglio se encuentra en contacto directo con los de los segmentos vecinos. A medida que el embrión crece en elongación, se van separando los ganglios y aparecen nervios de conexión entre los mismos. La cadena de ganglios conectados entre sí constituye el cordón nervioso ventral.

Andrew P. Kramer, colega nuestro, ha llegado a disecar el sistema nervioso en desarrollo de embriones de *Haementeria* e introdujo en sus neuronas microelectrodos y micropipetas. Ello le permitió registrar potenciales eléctricos de las neuronas embrionarias e inyectar sustancias fluorescentes en las mismas. Kramer ha observado que cuando los ganglios segmentarios aparecen en su forma globular, al principio, sus neuronas carecen de axones y dendritas, las prolongaciones características de las neuronas adultas. Las neuronas inmaduras están conectadas entre sí, de modo que el colorante inyectado puede pasar de una célula a otra. Más adelante, las neuronas desarrollan axones y desaparecen las primitivas conexiones intercelulares. Las neuronas embrionarias, sin embargo, carecen aún de la actividad eléctrica propia de la neurona adulta.

La actividad eléctrica se presenta en un estadio ulterior, cuando los axones en crecimiento alcanzan los nervios de conexión y los nervios segmentarios. Por el tiempo en que se asiste al cierre del cilindro corporal, el sistema nervioso goza ya de todas las propiedades que caracterizan al adulto. Las neuronas



EMBRIONES de la sanguijuela gigante *Haementeria*, fotografiados, en cuatro estadios sucesivos de su desarrollo, también por Sawyer, adscrito al grupo de los autores. En todas las microfotografías, la futura cabeza se encuentra en la parte superior. En general, la expansión de la placa germinal se acompaña de un progresivo alargamiento del eje longitudinal en relación a los ejes horizontal y vertical, transformando gradualmente el embrión primitivo ovoide en la sanguijuela joven de forma alargada y aplanada. En la parte superior izquierda, la fotografía corresponde a un embrión de nueve días (final del estadio 8). Se ha formado casi totalmente la placa germinal en la línea media ventral, excepto en la parte más distal, donde las bandas germinales no se han unido todavía. La parte frontal de la placa germinal está dividida ya en los grupos o bloques tisulares segmentarios. Las células de la placa germinal se han teñido de violeta para distinguirlas del vitelo amarillo. La fotografía de la parte superior derecha corresponde a un embrión de 11 días de edad (estadio 9). La placa germinal en expansión ha cubierto ya alrededor de un cuarto de la superficie ventral y los segmentos están separados por tabiques. En el tercio frontal de la placa germinal, cada segmento posee un ganglio completo en la línea media ventral; en el tercio medio, los segmentos tienen ganglios primordiales en varias fases de coalescencia, mientras que en el tercio inferior no se aprecian aún tales estructuras. La fotografía de la parte inferior izquierda corresponde a un embrión de 14 días de edad (estadio 10). La placa germinal ha cubierto ahora toda la superficie ventral y se han formado todos los ganglios segmentarios del cordón nervioso ventral, que están unidos por nervios de conexión. La ventosa posterior ha iniciado su constitución. En la parte inferior derecha se encuentra la fotografía de un embrión de 19 días (estadio 11). Los bordes izquierdo y derecho de la placa germinal en expansión se han unido en la línea media dorsal, cerrando el cilindro corporal. El intestino, que contiene aún el vitelo, se subdivide en múltiples compartimentos segmentarios, denominados ciegos; la ventosa posterior se distingue en forma de un disco tenue. Los embriones se muestran con distintas ampliaciones.

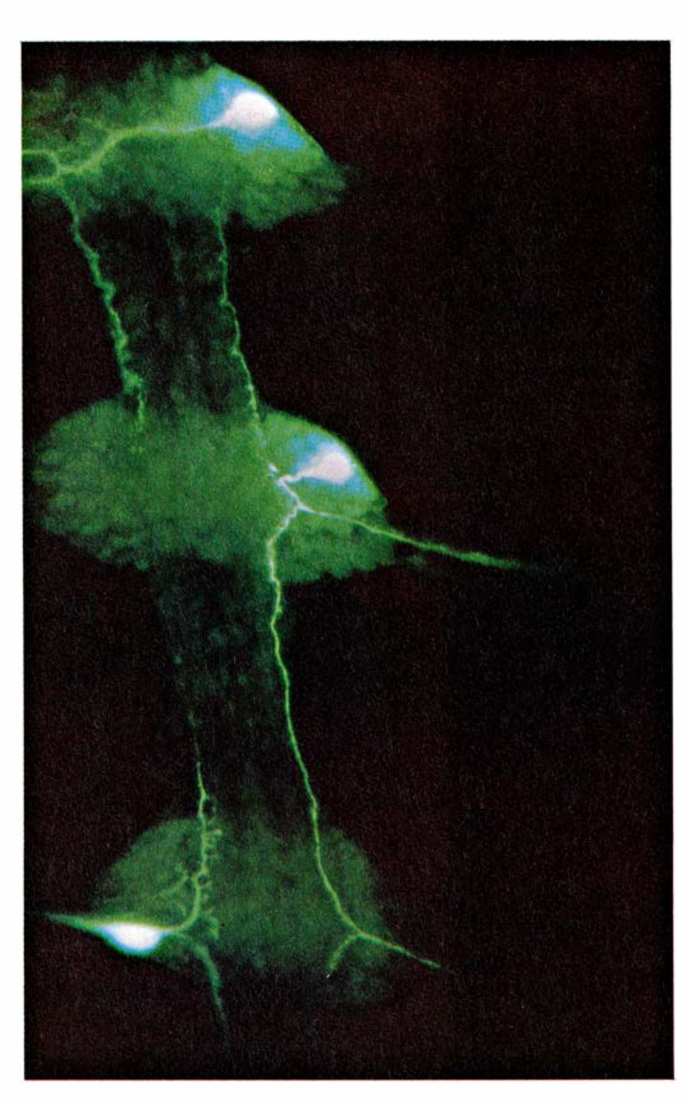
muestran entonces cambios de potencial parecidos a los que se observan en las neuronas maduras. Tal actividad eléctrica sugiere que las neuronas establecen conexiones sinápticas, a través de las cuales los impulsos nerviosos pasan de una célula a otra.

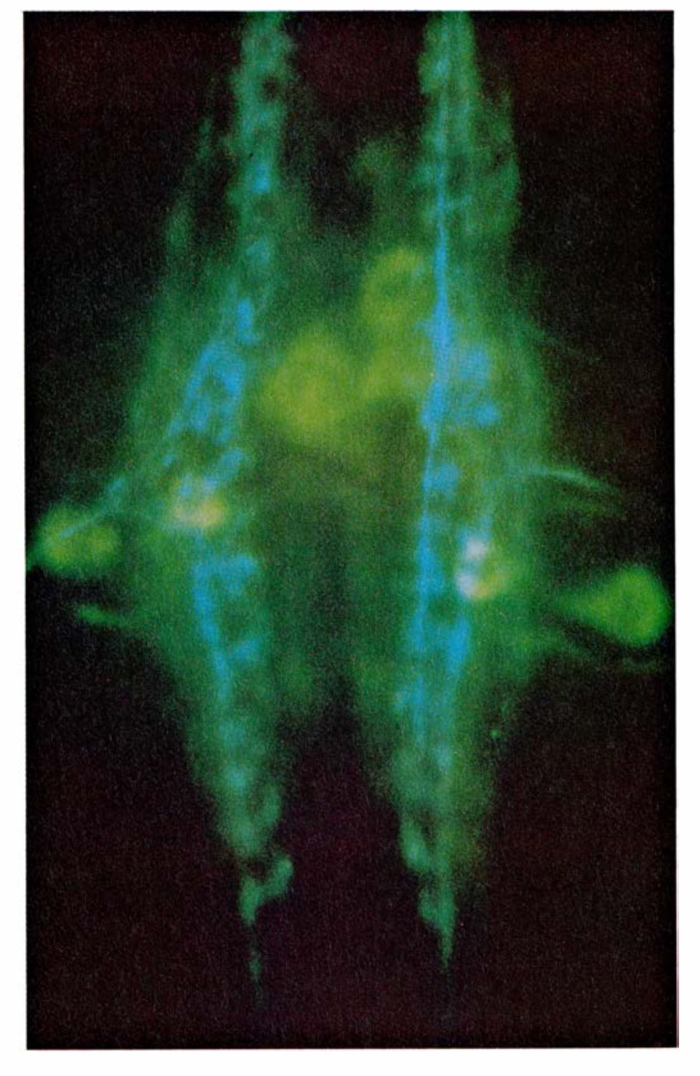
Puesto que el funcionamiento de una sinapsis depende de la liberación de ciertas sustancias, los denominados neurotransmisores, al tiempo del cierre corporal las neuronas embrionarias deben haber alcanzado la madurez suficiente para sintetizar tales sustancias. Los estudios realizados por nuestro colega Duncan K. Stuart han demostrado que así es, en efecto. Para determinar cómo se desarrolla la capacidad de síntesis del neurotransmisor acetilcolina, Stuart disecó los cordones nerviosos de embriones de sanguijuela de diferentes edades y los colocó en una disolución que contenía colina (precursor metabólico de la acetilcolina) marcada con tritio (isótopo radiactivo del hidrógeno). La velocidad de formación de acetilcolina marcada aportaba un índice cuantitativo de la concentración de enzimas que catalizaban la captación de colina y su conversión en acetilcolina.

Stuart observó que la capacidad de síntesis de acetilcolina era todavía muy baja cuando los ganglios tenían ya desarrollada su morfología y sus neuronas apenas comenzaban a emitir axones. La capacidad de síntesis se incrementaba rápidamente, sin embargo, cuando las neuronas enviaban axones a los nervios de conexión y segmentarios y producían potenciales de acción. Al tiempo del cierre corporal, la capacidad del cordón nervioso para sintetizar acetilcolina se había incrementado más de 25 veces sobre su valor inicial.

Stuart estudió otros dos neurotransmisores más: la dopamina y la serotonina. Si trataba los ganglios del organismo adulto con el colorante fluorescente ácido glioxílico, la sustancia se incorporaba fundamentalmente en las neuronas que contenían dopamina y serotonina, identificables luego al microscopio gracias a su fluorescencia verde. Expuso el sistema nervioso de embriones de distintas edades al ácido glioxílico; observó que las neuronas que contenían dopamina y serotonina empezaban a acumular el neurotransmisor aproximadamente al mismo tiempo que otras neuronas adquirían la capacidad de sintetizar acetilcolina.

Paralela a la maduración morfológi-





GANGLIOS SEGMENTARIOS del cordón nervioso de embriones de *Haementeria* en el estadio del cierre corporal. En cada uno de los tres ganglios de la izquierda se aprecia una neurona en la que se ha inyectado una sustancia fluorescente amarilla. La neurona del ganglio anterior (*arriba*) proyecta su axón en el nervio segmentario del lado opuesto del ganglio, y una rama en el nervio de conexión posterior que une el ganglio anterior con el medio. Las neuronas de los ganglios medio y posterior proyectan sus axones en el nervio segmentario del mismo lado, y ramas en los nervios de conexión anterior y posterior. La neurona del ganglio medio manda también una rama al nervio segmentario del ganglio posterior inmediato. El ganglio de la microfotografía de la derecha se ha tratado con ácido glioxílico, que induce fluorescencia específicamente en neuronas que contienen las sustancias neurotransmisoras dopamina y serotonina. La fluorescencia amarillo verdosa identifica siete células que contienen serotonina: tres células apareadas y una desapareada (*parte superior central*). La fluorescencia azul verdosa, más difusa, corresponde a los axones de dos pares de neuronas que contienen dopamina y cuyos cuerpos celulares se encuentran fuera del ganglio. La microfotografía de la izquierda es de Andrew P. Kramer; la de la derecha la tomó Duncan K. Stuart.

ca, bioquímica y eléctrica del sistema nervioso de la sanguijuela, tiene lugar la evolución de la conducta del embrión; se inicia ésta en un estadio primitivo y prosigue a través de una secuencia estereotipada de actos motores. Los análisis realizados por Kramer a partir de grabaciones en vídeo de embriones de *Haementeria* han puesto de manifiesto que la conducta del embrión progresa desde simples sacudidas irregulares hasta pautas de movimiento complejas. Los movimientos simples de los estadios iniciales quedan incorporados en los movimientos más elaborados que se observan en fases ulteriores. Los movimientos más complejos son, a su vez, parte de los hábitos motores de los animales jóvenes y adultos. Algunos de los movimientos embrionarios pueden desempeñar una función fisiológica en el momento en que aparecen. Coadyuvan a la circulación de nutrientes y a la salida del propio embrión a través de la membrana del huevo. Otros movimientos, sin embargo, no representan más que preludios de la conducta del adulto, concomitantes con la formación de conexiones funcionales entre los componentes del sistema nervioso en desarrollo.

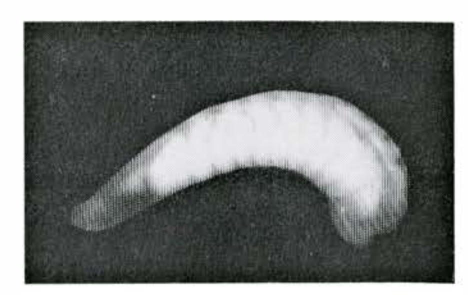
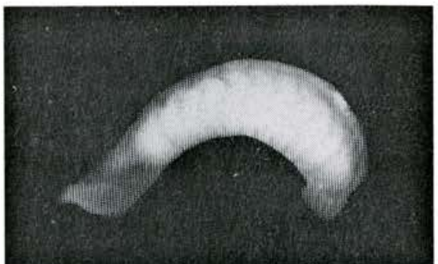
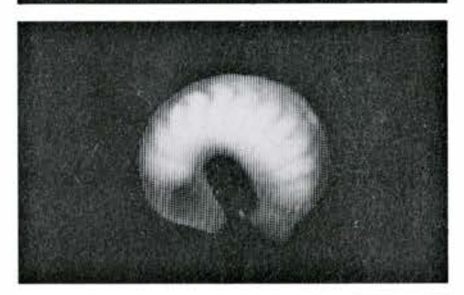
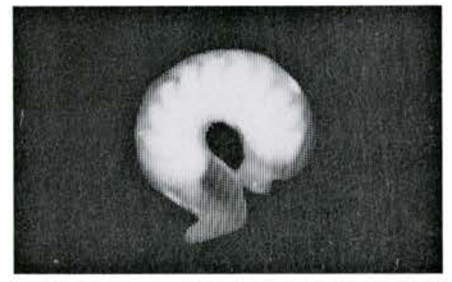
Los movimientos que prefiguran de forma patente la conducta del adulto empiezan con una onda de contracción peristáltica, que se observa cuando la placa germinal completa su coalescencia. Puesto que no hay sistema nervioso en este estadio del desarrollo, las contracciones peristálticas deben de representar un ritmo miogénico, es decir, de origen muscular. Los movimientos peristálticos dan como resultado la salida del embrión de la membrana del huevo. Ya fuera, los movimientos peristálticos del individuo disminuyen gradualmente, iniciándose movimientos de incurvación lateral intermitentes, que se deben a la contracción de los músculos longitudinales de un lado y a la distensión de los del lado opuesto. En ese momento, el cordón nervioso ventral del embrión está suficientemente desarrollado para ejecutar movimientos neurogénicos.

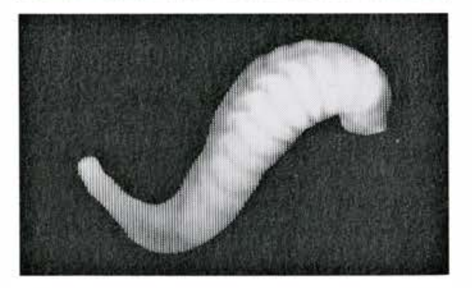
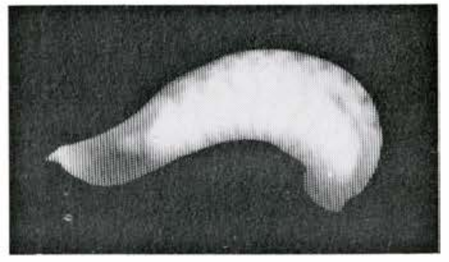
Después del cierre corporal, los movimientos de flexión lateral dan paso a un ciclo motor más complicado, preludio del reptar característico de la sanguijuela joven [*véase la ilustración de esta página*]. Ulteriormente, cuando se han desarrollado las ventosas frontal y caudal, aparece la reptación característica del joven anélido. Llegada a esta fase, la diminuta sanguijuela coordina ya la contracción de sus músculos longitudinales con el ritmo característico de los movimientos de natación. En consecuencia, la joven *Haementeria* está dotada para buscar su alimento por reptación o natación, dirigiéndose hacia el hipotético huésped.

Nos preguntábamos en el comienzo del artículo por los orígenes embrionarios de las células que componen el sistema nervioso de la sanguijuela. Whitman y sus discípulos establecieron, por observación microscópica directa de embriones, el linaje celular, a partir del huevo, de las células *A, B* y *C* de los teloblastos y de las células progenitoras que forman las cintillas. Con el incremento continuo del número de células en estadios ulteriores, la observación directa resulta un método excesivamente complicado para seguir el destino de las distintas células. Por ello, y a fin de disponer de una genealogía completa de las células del sistema nervioso en desarrollo de las sanguijuelas, utilizamos una nueva técnica de marcaje.

En la nueva técnica se inyecta una molécula marcada, mediante una micropipeta, en una determinada célula del embrión primitivo. Se deja que continúe el desarrollo embrionario hasta una fase ulterior y se examina entonces la distribución de la sustancia marcada en los tejidos. La sustancia seleccionada para tal experimento ha de satisfacer tres requisitos: el desarrollo embrionario tiene que continuar normalmente después de la inyección, la sustancia debe permanecer intacta (evitándose la dilución excesiva) y, por último, dicho marcador no debe atravesar las uniones que ponen en comunicación células embrionarias, de modo que pase exclusivamente a células que sean descendientes directas de la inyectada.

Uno de los marcadores elegidos fue la rodamina, una sustancia con fluorescencia roja; le acoplamos un polipéptido transportador de secuencia conocida. En colaboración con Janis D. Young sintetizamos el polipéptido a partir de aminoácidos cuya configuración se apartaba de la natural, al objeto de protegerlo de la acción degradadora

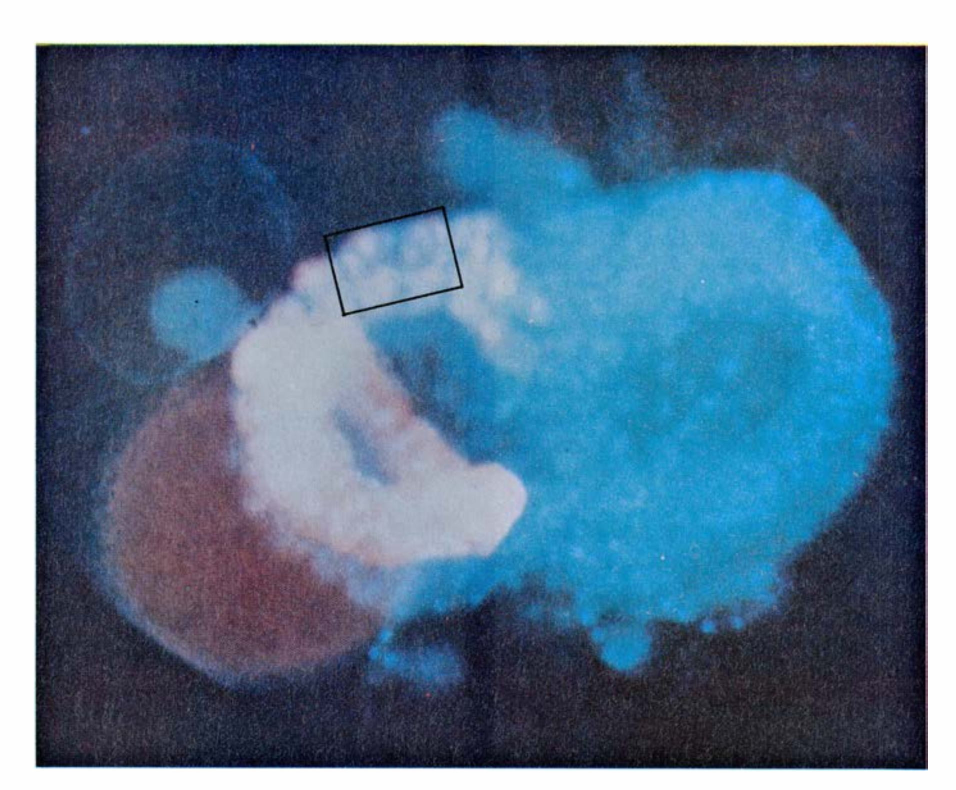





MOVIMIENTO DE PRE-REPTACION de un embrión de *Haementeria* en el estadio del cierre corporal, registrado en vídeo por Kramer. El tiempo total de la secuencia dura 25 segundos. En un estadio posterior aparece la reptación característica del adulto. En éste, los movimientos forman un ciclo que se inicia fijando la ventosa posterior en una base sólida y alargando el cuerpo. Luego, se fija la ventosa anterior y se libera la posterior. El cuerpo se encoge ahora, curvándose hacia abajo y hacia dentro en sus extremos, tras lo cual la ventosa posterior vuelve a fijarse en una posición más avanzada y la ventosa anterior se desprende. El ciclo se completa cuando el cuerpo se endereza y distiende.

de los enzimas digestivos. El complejo rodamina-polipéptido resultó particularmente útil para visualizar el desarrollo de la placa germinal. ¿Cómo se procede? Se inyecta un teloblasto o una célula precursora de un teloblasto junto con el complejo rodamina-polipéptido; luego se fija el embrión, se le elimina el vitelo nutritivo y se trata con una sustancia de fluorescencia azul, específica para el ADN. El resultado es la tinción con fluorescencia azul de todos los núcleos celulares y, con fluorescencia roja, del citoplasma de las células descendientes del teloblasto inyectado.

En uno de tales experimentos, llevado a cabo por nuestro colega Saul L. Zackson, en la célula *OPQ* derecha (precursora de los teloblastos derechos *O*, *P* y *Q*) se inyectó el complejo rodamina-polipéptido en el estadio 6, y se examinó el embrión en el estadio 8 [*véase la ilustración de la página 87*]. Las tres células de gran tamaño coloreadas de rojo que aparecen en la micrografía son los teloblastos derechos *O*, *P* y *Q*. De cada uno de ellos parte una cintilla de células progenitoras hijas que se proyecta hacia la banda germinal derecha. Los núcleos de las células pertenecientes a cada cintilla pueden reconocerse porque se encuentran en las células rojas. Si prestamos atención a los núcleos de la cintilla, advertiremos cómo hay algunos en proceso de división.





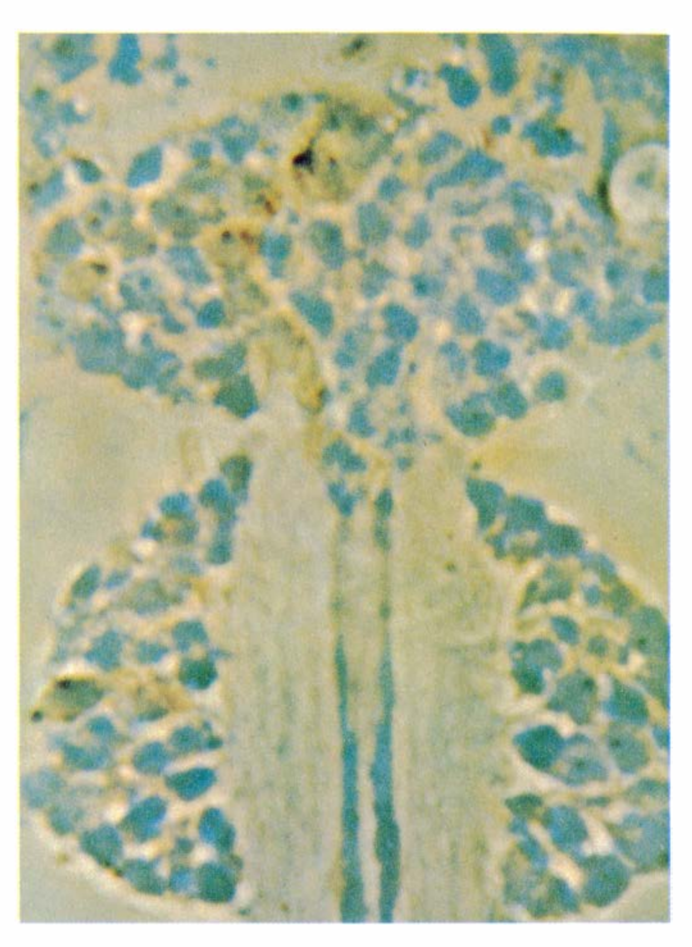
FORMACION DEL MESODERMO, la capa media de tejido germinal del embrión de sanguijuela, observada con una nueva técnica de marcaje utilizada por los autores. En los teloblastos *M* del embrión de la microfotografía superior se inyectó, en el estadio 5, el complejo rodamina-polipéptido, que muestra fluorescencia roja. La preparación se hizo en el estadio 8, tras fijar y eliminar el vitelo y tratar el tejido con una sustancia de fluorescencia azul. Se observa una cintilla de células progenitoras de color rojo proyectándose hacia arriba, a partir de un teloblasto *M* rojo de gran tamaño. La cintilla se incurva bruscamente en la superficie del embrión antes de unirse a la banda germinal izquierda en forma de media luna. El rectángulo negro marca dos grupos adyacentes de células progenitoras mesodérmicas, reproducidas a mayor aumento en la imagen inferior. Los grupos son isomorfos (morfológica y topológicamente similares), lo que sugiere que el desarrollo inicial del mesodermo tiene lugar a través de una secuencia estereotipada de divisiones celulares. Las microfotografías fueron tomadas por Saul L. Zackson, de Berkeley.

En otro experimento se examinó, también en el estadio 8, un embrión en cuyo teloblasto *M* izquierdo se había inyectado el complejo rodamina-polipéptido en el estadio 5 [*véase la ilustración de esta misma página*]. Aquí, el teloblasto *M*, rojo y de gran tamaño, está situado profundamente (como cabe suponer de una célula que da lugar al mesodermo), y está desenfocado. A partir de él, una cintilla de células progenitoras rojas se proyecta hacia arriba; al alcanzar la superficie, la cintilla se curva bruscamente hacia la izquierda y se une a la banda germinal izquierda en forma de media luna. En su extremo frontal, la cintilla mesodérmica está subdividida en agrupaciones de células marcadas. La subdivisión constituye el primer signo de segmentación del tejido mesodérmico, que empieza incluso antes de que las bandas germinales izquierda y derecha establezcan coalescencia.

La ampliación fotográfica del extremo frontal de la banda germinal de este embrión muestra que los grupos de células adyacentes tienen un aspecto semejante. El parecido, que recibe el nombre técnico de isomorfismo, sugiere que el desarrollo inicial del mesodermo de cada segmento corporal se lleva a cabo a través de una secuencia de divisiones celulares estereotipadas. Experimentos más recientes, llevados a cabo por Zackson, han puesto de manifiesto que cada grupo de células procede de una célula progenitora única que, a su vez, deriva de un teloblasto *M*. Cada grupo o bloque es, a buen seguro, precursor de todo el tejido mesodérmico que queda a un lado de la línea media del segmento. Por tanto, cada uno de los 32 segmentos corporales mesodérmicos se origina como un par bilateral de células progenitoras derivadas de un teloblasto *M*.

El otro marcador que utilizamos para





OTRA TECNICA DE MARCAJE empleada por los autores para establecer la genealogía celular en embriones de sanguijuela. En los dos embriones, vistos ventralmente en la microfotografía de la izquierda, se inyectó el enzima peroxidasa de rábano en el estadio 6. En el estadio 10, los embriones se trataron con sustancias químicas que teñían de un color oscuro las células inyectadas y sus descendientes. En el embrión del lado izquierdo de la microfotografía, el enzima se inyectó en el teloblasto izquierdo *OPQ* y, en el embrión de la derecha, el enzima se inyectó en el teloblasto izquierdo *N*. La microfotografía de la derecha muestra, a mayor aumento, una sección horizontal ligeramente oblicua de un ganglio segmentario del cordón nervioso de un embrión en el que se inyectó una célula *OPQ* de forma similar al embrión izquierdo de la primera microfotografía. Las neuronas que contienen la peroxidasa de rábano destacan por su color oscuro sobre un fondo azul. Parte de la sección atraviesa el tracto nervioso central, mostrando los axones que cursan longitudinalmente.

establecer la genealogía de las células embrionarias fue el enzima peroxidasa de rábano. A mediados de la década de 1970, Kenneth J. Muller y Uel J. McMahan II, de la Facultad de Medicina de Harvard, fueron los primeros en desarrollar el procedimiento de inyección de peroxidasa de rábano en las neuronas del sistema nervioso de sanguijuelas adultas. Exponiendo luego el tejido a peróxido de hidrógeno y bencidina, la totalidad de la célula nerviosa inyectada, incluidas las extensiones más diminutas del axón y las dendritas, adquiría un color oscuro.

Inyectamos peroxidasa de rábano en embriones que se encontraban en el estadio 6, cuando la placa germinal se extiende hasta cubrir prácticamente la totalidad de la superficie ventral y cuando están presentes ya los 32 ganglios segmentarios del cordón nervioso [*véase la ilustración de esta misma página*]. En un embrión particular, se introdujo el enzima en un teloblasto *N* izquierdo. El marcador apareció en estructuras segmentarias repetidas, cuyo tamaño, forma y posición indicaban que se trataba de mitades de los ganglios del lado izquierdo del cordón nervioso ventral. De ello se deducía que el par de teloblastos *N* era el progenitor de una parte sustancial del sistema nervioso de la sanguijuela. Sin embargo, los ganglios no quedaban teñidos de manera uniforme. En cada mitad del ganglio, el marcador formaba una cinta fina longitudinal, próxima a la línea media ventral, y dos transversales, que se extendían desde ella lateralmente. La presencia de tejido no teñido entre las dos cintas transversales en cada ganglio indicaba que algunas neuronas del sistema nervioso segmentario tenían un origen distinto al del teloblasto *N*.

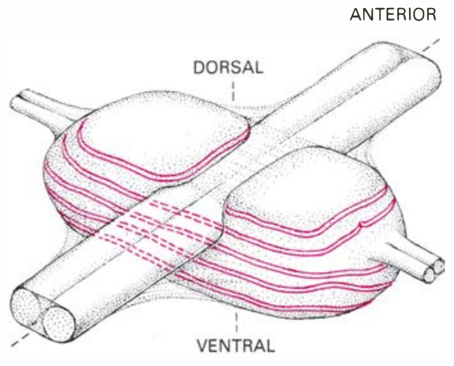
Un segundo embrión, en el que se marcó con peroxidasa de rábano la célula *OPQ* izquierda, demostró que las neuronas restantes se originaban en los teloblastos *O*, *P* y *Q*. Si bien la mayor parte del marcador teñía el ectodermo de la placa germinal izquierda, había también algunas manchas, repetidas segmentariamente, cuya posición próxima a la línea media ventral señalaba que marcaban neuronas de ganglios segmentarios. Los patrones de marcaje de ambos embriones resultaron complementarios: los fragmentos de tejido nervioso teñidos después de la inyección de peroxidasa de rábano en la célula *OPQ* eran los que quedaban sin teñir cuando el enzima se inyectaba en el teloblasto *N*.

La distribución celular de la peroxidasa de rábano se determinó con mayor precisión examinando cortes seriados de embriones. Las secciones realizadas a través del cordón nervioso de embriones con teloblastos *N* inyectados indicaron que la tenue cinta longitudinal teñida correspondía a un bloque celular de la porción de ganglio más próxima a la línea media; las dos cintas transversas correspondían a un bloque anterior y otro posterior de neuronas en la parte ganglionar más alejada de la línea media. Las secciones realizadas a través del cordón nervioso de embriones cuya célula *OPQ* se había inyectado mostraron que las agrupaciones o bloques celulares teñidos correspondían a bloques no teñidos cuando se marcaba el teloblasto *N*.

DORSAL
ANTERIOR
POSTERIOR
CELULAS MUSCULARES
CORDON NERVIOSO CENTRAL
CELULAS GLIALES
VENTRAL

Experimentos ulteriores, en los que se inyectaba separadamente peroxidasa de rábano en los teloblastos *O, P* y *Q* (en lugar de en su precursor común) pusieron de manifiesto que los descendientes de cada teloblasto formaban bloques neuronales característicos en cada ganglio segmentario. El tamaño y posición de cada grupo no variaba de un segmento a otro, ni entre distintas muestras. Esta observación sugirió que cada uno de los cuatro pares de teloblastos daba origen a una familia neuronal distinta. Además, vista la posición invariable de las neuronas identificadas en los ganglios segmentarios, cada neurona sería el descendiente lineal de un determinado teloblasto.

Los análisis genealógicos de este tipo muestran que la neurogénesis de la sanguijuela está altamente predeterminada, en el sentido de que un teloblasto dado siempre resultará en una misma parte del sistema nervioso. Otro aspecto de la predeterminación lo constituye el hecho de que, cuando un teloblasto muere o tiene un funcionamiento defectuoso, su papel en el desarrollo no es asumido por ninguna otra célula. La destrucción de un teloblasto en el



MAPAS GENEALOGICOS que identifican el linaje de las células de un ganglio segmentario de un embrión de sanguijuela. El esquema de la izquierda comprende cinco secciones horizontales de un ganglio segmentario situado en la porción media de un embrión de *Helobdella* en el estadio 10. Un esquema del ganglio intacto se muestra en la parte superior. Los dos pares de formaciones alargadas sombreadas en el centro de la segunda sección, empezando por arriba, son células musculares situadas en el tracto nervioso longitudinal. Descienden del par de teloblastos *M*. En la sección central, se observan dos células gliales que descienden de teloblastos *N*. Las formas redondeadas que se observan en cada sección no corresponden a células reales, pero representan aproximadamente el tamaño, número y disposición de las neuronas de cada corte. En la mitad izquierda del ganglio, las zonas coloreadas contienen células descendientes del teloblasto *Q* izquierdo. En la mitad derecha del ganglio, las zonas coloreadas portan descendientes del teloblasto *N* derecho. Las porciones no coloreadas contienen descendientes de los teloblastos *O* y *P*.

embrión primitivo conduce a la aparición de malformaciones características. Seth S. Blair nos ayudó a analizar las consecuencias neuroanatómicas que se derivaban de la destrucción de determinados teloblastos, o de sus células precursoras, tras la inyección de enzimas tóxicos.

En línea con nuestros estudios sobre la genealogía celular, observamos que la destrucción de un teloblasto *N* en el estadio 6 daba lugar a la formación de un sistema nervioso embrionario cuyos ganglios segmentarios mostraban déficit anatómicos en el lado alterado. Sin embargo, el patrón de las alteraciones no era tan regular como cabría esperar. Los hemiganglios del lado alterado variaban notablemente de tamaño en un mismo organismo, observándose, desde un número casi normal de células, hasta la total ausencia de neuronas segmentarias.

La mitad ganglionar, que era casi normal en cuanto a su tamaño se refiere, incluía las células que normalmente proceden de los teloblastos *O, P* y *Q* del mismo lado. También había en ellos algunas células derivadas de teloblastos *N* intactos del otro lado; estas últimas células habrían cruzado de forma anormal la línea media de la placa germinal durante la formación de los primordios ganglionares. De la ausencia casi total de algunos hemiganglios cabía deducir que el fenómeno anterior estaba sometido a notables fluctuaciones. Se puso de manifiesto que las células derivadas de los teloblastos *N* cumplían alguna misión organizativa de las células derivadas de las *O, P* y *Q*, a la hora de localizar éstas los primordios ganglionares.

La destrucción de un teloblasto *M* daba lugar, según lo esperado, a un embrión carente de todas las estructuras mesodérmicas del lado alterado. Además, el lado afectado carecía de sistema nervioso. Los teloblastos *N, O, P* y *Q* normales producían sus cintillas de células progenitoras; mas, en ausencia de grupos tisulares mesodérmicos, las células precursoras del sistema nervioso no se organizaban en primordios ganglionares. Los déficit causados por la destrucción de teloblastos denotan la importancia de las interacciones entre células de diferente genealogía en el desarrollo del sistema nervioso de la sanguijuela.

Así pues, ¿cómo surgen las neuronas del sistema nervioso de la sanguijuela y cómo se forman sus conexiones específicas durante el desarrollo? De lo expuesto anteriormente resulta evidente que las respuestas dadas a tan importantes cuestiones han sido más descriptivas que explicatorias. No obstante, estamos ya en condiciones de formular una serie de cuestiones más concretas sobre los mecanismos del desarrollo del sistema nervioso. ¿Está predeterminada la sucesión precisa de planos de división alternantes, primero durante la embriogénesis temprana y, después, durante la formación de los segmentos mesodérmicos? Si así es, la predeterminación probablemente refleje la replicación y situación de los elementos del citoesqueleto de las células embrionarias, puesto que es la orientación del citoesqueleto la que gobierna el plano de división celular.

Una cuestión relacionada se refiere al papel del teloplasma que pasa únicamente a la célula *D*. Tal como se explicó antes, la parte ventral del teloplasma pasa a la célula *DM*, y, por tanto, a las células precursoras del mesodermo, mientras que la parte dorsal pasa a la célula *DNOPQ* y a las células precursoras del ectodermo. La mayor parte del teloplasma se reparte entre las células descendientes de los teloblastos. ¿Qué mecanismo asegura que sean exactamente 32, y no 31 o 33, las células descendientes del teloblasto *M* que encuentran una agrupación segmentaria de tejido mesodérmico a cada lado del embrión? ¿Cómo se reúnen las células descendientes de los teloblastos *N, O, P* y *Q* precursores del ectodermo, situadas en una agrupación tisular de este tipo, para formar las neuronas de los primordios ganglionares? ¿Puede atribuirse la maduración morfológica, eléctrica y química de las neuronas a la interacción de sus axones en desarrollo con las células o tejidos efectores? ¿Acaso el desarrollo de las neuronas motoras y de sus conexiones, que constituyen la base del perfeccionamiento gradual de la reptación y natación, requiere retroalimentación funcional, o son simplemente la consecuencia de mecanismos automáticos que surgen durante la embriogénesis?

Finalmente, ¿cómo se determinan las características definitivas de una neurona a partir de su línea genealógica? ¿Recibe cada neurona un lote particular de determinantes intracelulares, que se repartirán, a partir de la célula precursora, entre las células hijas en divisiones sucesivas? ¿O quizá las características de la célula vienen establecidas por su posición en el embrión y, en consecuencia, por los patrones de división a través de los que se han originado? Esperemos que las técnicas experimentales ya disponibles resuelvan pronto algunos de estos interrogantes.

Psicología de las preferencias

Cuando elegimos no siempre lo hacemos objetivamente. Estudios experimentales demuestran que tales faltas de objetividad tienden a seguir patrones regulares que admiten una descripción matemática

Daniel Kahneman y Amos Tversky

Imagine que se dirige al teatro con un par de localidades por las que ha pagado 2000 pesetas. Al llegar, descubre que ha perdido las entradas. ¿Volvería a gastarse el mismo dinero por unas nuevas localidades? Ahora imagínese yendo al mismo teatro sin haber comprado las localidades. Y al llegar se da cuenta de que ha perdido 2000 pesetas en metálico. ¿Compraría entradas para esa obra?

En términos objetivos, las dos situaciones son idénticas: en ambos casos ha sufrido un empobrecimiento de 2000 pesetas y debe tomar la decisión de pagar otro tanto para asistir a la representación, o de no hacerlo y perderse la obra. Sin embargo, la mayoría de las personas a las que se presenta tal situación afirman que estarían más dispuestas a comprar las entradas si hubiesen perdido el dinero que si hubiesen perdido los billetes de acceso al teatro. Una posible interpretación de las diferencias entre ambas situaciones es que la misma pérdida puede asignarse a diferentes partidas en una especie de cuenta corriente mental. La pérdida de 2000 pesetas en efectivo entra en una partida no relacionada con la obra teatral. Por ello, tal pérdida incide poco en la decisión de comprar nuevas entradas. Por el contrario, el costo de los billetes perdidos se carga en la cuenta de la representación teatral. Y es difícil aceptar el costo doble o inesperado del precio de las localidades.

Recientes investigaciones en el campo de la psicología de las preferencias han mostrado algunas curiosas discrepancias entre las razones subjetivas y objetivas de la decisión. Por ejemplo, el riesgo de una pérdida influye más en la decisión que la posibilidad de una ganancia equivalente. Las personas, en su mayoría, son también muy sensibles a la diferencia entre certeza y alta probabilidad, y relativamente insensibles a gradaciones intermedias de probabilidad. El sentimiento asociado a una pérdida que tiene lugar como consecuencia de una acción tiende a ser más intenso que el asociado a una pérdida ocasionada por una omisión, o por una oportunidad no aprovechada. Estas, y otras observaciones de carácter similar, contribuyen a comprender mejor el modo en que la gente toma decisiones, y a aclarar algunos enigmas de la elección racional.

El origen de la psicología de las preferencias puede remontarse a un ensayo publicado en 1738 por el matemático Daniel Bernouilli. En aquel trabajo, Bernouilli se refirió a una característica muy difundida de las preferencias humanas: la aversión al riesgo. Para hacerse idea de la misma, imagine que se le da a elegir entre dos opciones: la primera es una ganancia segura de 8000 pesetas. La segunda es un juego, evidentemente arriesgado, que se basa en la probabilidad del 85 por ciento de ganar 10.000 pesetas y el 15 por ciento de no ganar nada.

La mayoría de las personas a las que se presenta esta elección prefieren la ganancia segura a la probabilidad, a pesar de que asumir el juego de azar ofrece una mayor expectativa monetaria que optar por el resultado seguro. La expectativa monetaria de un juego es la suma de sus posibles resultados multiplicados por sus probabilidades respectivas. La expectativa monetaria del juego presentado en este caso es igual a 10.000 pesetas por su probabilidad (0,85), más cero pesetas por la suya (0,15), o, lo que es lo mismo, 8500 pesetas. La expectativa monetaria expresa el valor promedio esperado; si uno hubiese de jugar este juego muchas veces, el promedio de ganancia sería de 8500 pesetas por partida. La expectativa monetaria de la ganancia segura es 8000 multiplicado por 1 (certeza), o un total de 8000 pesetas. Una elección es adversa al riesgo si se prefiere un resultado seguro a un juego que posea una expectativa monetaria igual o mayor. Por otra parte, una elección es favorable al riesgo si se rechaza un resultado seguro a cambio de un juego que posea una expectativa monetaria igual o menor.

La hipótesis de que la gente suele realizar elecciones adversas al riesgo ha sido ampliamente aceptada por los economistas, quienes suponen que un consumidor o un empresario van a preferir un negocio u operación arriesgados, frente a algo seguro, sólo cuando la expectativa monetaria de esa operación sea lo suficientemente alta como para compensarles por asumir ese riesgo. Sin embargo, los estudios psicológicos indican que son muy frecuentes las elecciones favorables al riesgo cuando ha de elegirse entre una pérdida segura y una alta probabilidad de una pérdida mayor.

Para hacerse una mejor idea de las preferencias favorables al riesgo, imagine que se ve obligado a elegir entre una pérdida segura de 8000 pesetas y el riesgo que supone la probabilidad del 85 por ciento de perder 10.000 pesetas y el 15 por ciento de no perder nada. Frente a esta elección, una gran mayoría de personas prefieren la probabilidad a la pérdida segura. A pesar de todo, la expectativa monetaria del juego (−8500) es peor que la de la pérdida segura (−8000). El que exista una preferencia mayoritaria constituye entonces un ejemplo de elección favorable al riesgo.

La diferencia entre las elecciones mayoritarias que se observan en estos dos problemas sugiere que las preferencias adversas al riesgo corresponden a elecciones entre ganancias, mientras que las preferencias favorables al riesgo corresponden a elecciones entre pérdidas. Se ha observado este esquema tanto en respuestas a preguntas hipotéticas como en problemas reales de decisión, donde se pagaba a la gente en función de sus elecciones. El mismo modelo rige para problemas hipotéticos no relacionados con resultados pecuniarios, en casos tales como, por ejemplo, la dura-

ción del dolor, el número de vidas que pueden perderse en una epidemia o que se pueden salvar por intervención médica.

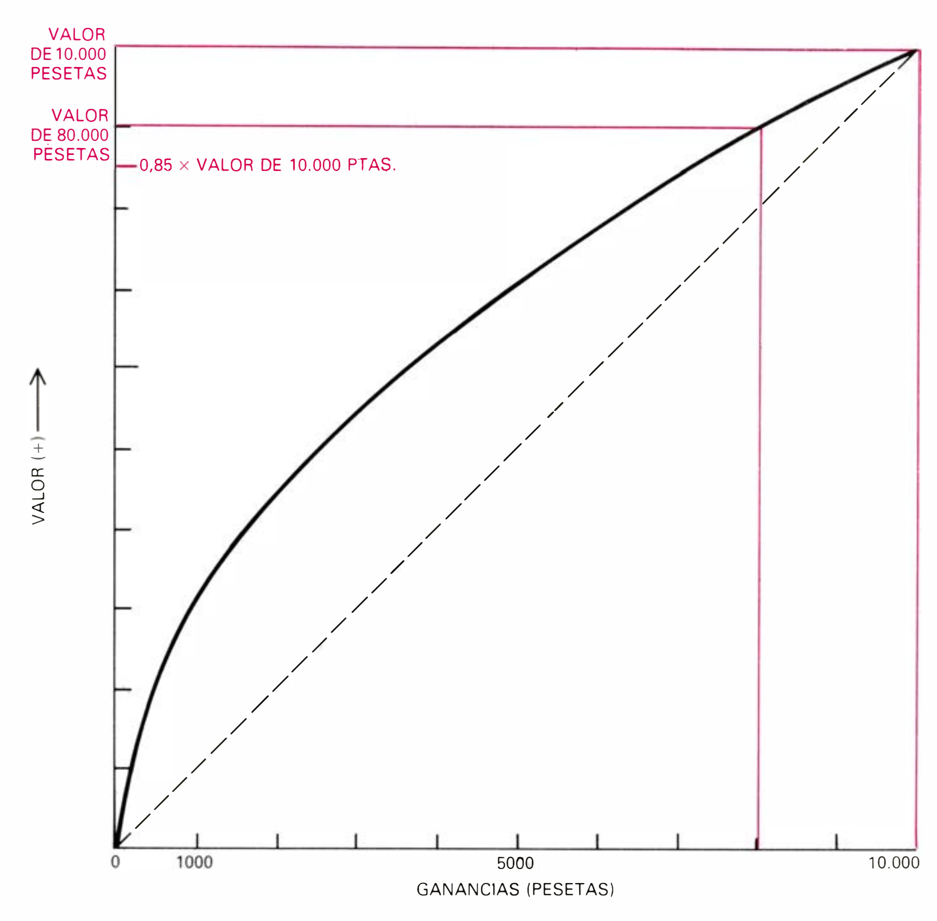
Cuando la gente efectúa elecciones adversas o favorables al riesgo, renuncia a la opción con mayor expectativa monetaria. Para explicar este hecho, Bernouilli sustituyó el criterio objetivo de la expectativa monetaria por el subjetivo de la utilidad esperada. De acuerdo con la teoría de la utilidad esperada, cada resultado de una elección da origen a un cierto grado de bienestar o de utilidad. La utilidad de una perspectiva arriesgada es la suma ponderada de las utilidades de sus resultados, ponderación que se efectúa a través de las probabilidades de esos resultados, multiplicándolas por las utilidades respectivas. La idea central es que la utilidad no sigue una función lineal de la cantidad de dinero: la utilidad de una ganancia de 200.000 pesetas supone menos del doble de la utilidad de una ganancia de 100.000 pesetas. Por tanto, la perspectiva con mayor expectativa monetaria no tiene necesariamente la mayor utilidad esperada. Se supone que el sujeto que ha de decidir selecciona la opción que posea la mayor utilidad, aunque tal opción pueda, o no, mantener la mayor expectativa monetaria.

¿Cómo identifica la gente los resultados de sus decisiones? De acuerdo con Bernouilli, se espera generalmente que la utilidad se relacione con estados de riqueza. Nosotros abandonamos esta tradición y preferimos analizar las elecciones en términos de cambios de riqueza. Del análisis clásico se deduce que las preferencias reflejan un punto de vista de las opciones en conjunto. Nosotros, por el contrario, proponemos que las personas adoptan normalmente un punto de vista limitado de los resultados de sus elecciones o decisiones: es decir, identifican las consecuencias como ganancias o pérdidas en relación con un punto neutro. Esta forma de contabilidad mental puede conducir a elecciones contradictorias, puesto que la misma consecuencia objetiva se puede valorar de muchas maneras distintas.

Para explicar adecuadamente las elecciones proponemos una función que relaciona un valor subjetivo con cualquier cantidad que se pueda ganar o perder. La llamaremos función de valor. Del mismo modo que en la teoría clásica de la utilidad, la aversión o la inclinación al riesgo pueden quedar explicadas por la curvatura de la función, que relaciona valores subjetivos con resultados objetivos. Consideremos primero la función de valor para el caso de elecciones entre ganancias. Suponemos que éstas tienen un valor positivo, y que una ganancia cero, que meramente mantiene el status quo, tiene un valor subjetivo de cero. Proponemos, también, que la función de valor para las ganancias es cóncava hacia abajo, de tal modo que cada unidad ganada añade menos valor que la precedente. La concavidad de la función de valor está de acuerdo con la creencia de sentido común de que la diferencia entre una ganancia de 10.000 pesetas y una de 20.000 es más significativa que la diferencia entre una ganancia de 110.000 y una de 120.000. Esta forma de la función de valor apoya la noción de aversión al riesgo.

Consideremos ahora la elección entre una ganancia segura y un juego en el que se ofrece cierta probabilidad de una ganancia mucho mayor y cierta probabilidad de no ganar nada en absoluto. La ventaja de la ganancia mayor sobre la segura puede observarse en la región plana de la función, donde los incrementos de dinero producen pequeños incrementos de valor. En cambio, la primacía del resultado seguro sobre la ganancia nula se evalúa en la región más inclinada de la función, en la que cada cantidad sumada se corresponde con una gran diferencia de valor. Puesto que el valor no es lineal, el resultado seguro corresponde a la máxima ganancia en términos de valor, pero no a la máxima en términos de dinero. La forma de la función de valor favorece la preferencia por los resultados seguros, en contra de las perspectivas arriesgadas, preferencia evidentemente adversa al riesgo [*véase la ilustración de esta misma página*].

Para explicar las frecuentes preferencias favorables al riesgo en las elecciones entre pérdidas, suponemos que la función de valor es convexa, de tal modo que cada cantidad perdida provoca menor cambio de valor que la precedente. Esta proposición es compatible



FUNCION DE VALOR CONCAVA, que ayuda a explicar las preferencias entre posibles ganancias. El valor representa el atractivo subjetivo de una ganancia (o una pérdida). La función asigna un valor a cada ganancia posible. Las ganancias se indican en el eje de abscisas; los valores, en el de ordenadas. Al status quo se le asigna valor cero. (Las unidades de valor son arbitrarias.) La función cóncava se hace progresivamente plana conforme crece la cantidad de ganancia. Cuando la gente elige entre varias ganancias, es frecuente observar preferencias adversas al riesgo. En una elección adversa al riesgo se prefiere una ganancia segura a un juego con expectativa monetaria mayor. La expectativa monetaria de un juego es la suma de sus resultados multiplicados por sus probabilidades. Por lo tanto, la probabilidad del 85 por ciento de ganar 10.000 pesetas tiene una expectativa monetaria de 0,85 multiplicado por 10.000, o sea 85.000 pesetas. Y es mayor que la expectativa monetaria de una ganancia segura de 8000 pesetas (8000 por la certeza, 1). El valor de la ganancia segura, por el contrario, es mayor que el del juego, que es 0,85 por el valor de 10.000 pesetas. La función de valor predice que se preferirá la ganancia segura al juego.

con la impresión común de que la diferencia entre la pérdida de 10.000 pesetas y la de 20.000 parece más significativa que la diferencia entre la pérdida de 110.000 y la de 120.000. La convexidad de la función de valor para las pérdidas favorece la preferencia por un juego favorable al riesgo, en lugar de la pérdida segura. La ventaja del mejor resultado (no perder nada) sobre la pérdida segura se evalúa en la parte más inclinada de la función de valor; por otra parte, la ventaja de la pérdida segura sobre el peor resultado del juego se observa en la región más plana. Como resultado, la pérdida segura está relativamente más cercana al resultado peor del juego en la escala de valor que en la escala de dinero [*véase la ilustración de esta misma la página*].

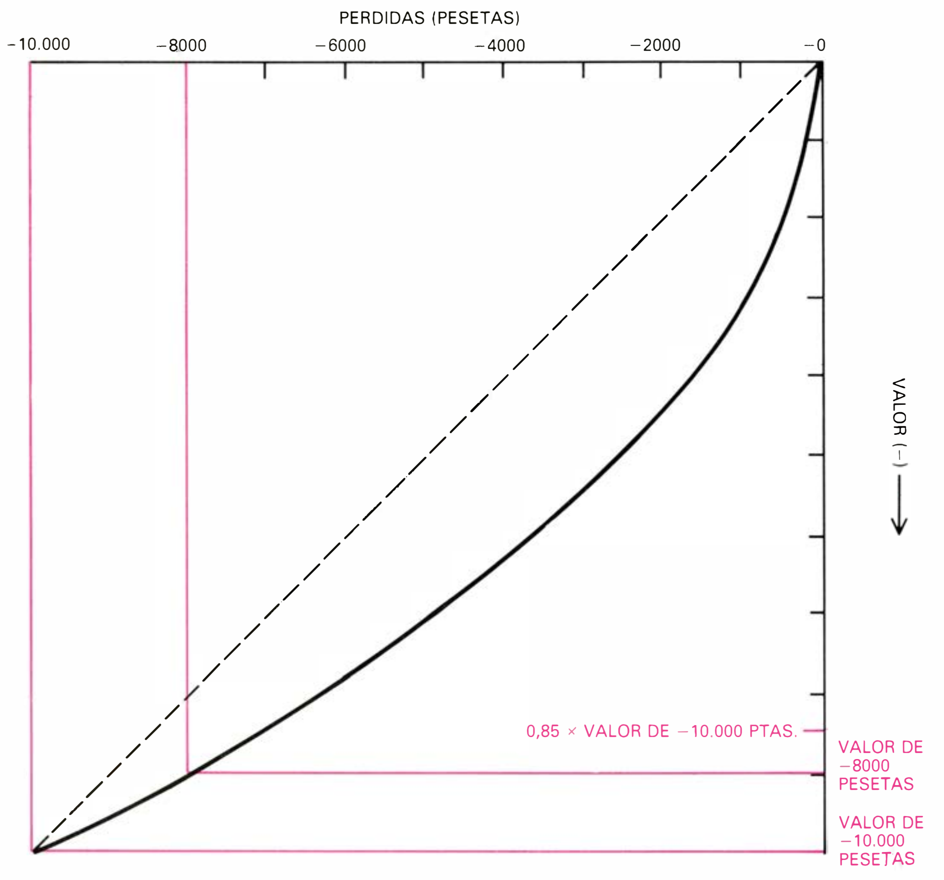
Podemos inferir las propiedades de la función de valor a partir de las elecciones observadas. Para comprender cómo funciona este procedimiento de construcción de la función, imagine que se le ofrece una elección entre ganar una cierta cantidad en metálico o la probabilidad del 50 por ciento de ganar 10.000 pesetas. ¿Qué cantidad de dinero hará tan atractivo al premio en metálico como el juego? Es frecuente que la cantidad sea de 3500 pesetas. Si halla tan atractivas las 3500 pesetas como el juego, el valor que tiene para usted esa cantidad iguala el valor que tiene para usted el juego. A partir de la presuposición que hemos presentado antes, el valor del juego es 1/2 del valor de 10.000 pesetas, dado que la probabilidad de ganar 10.000 pesetas es 1/2, y el valor del otro resultado (no ganar nada) se supone que es cero. Consecuentemente, emparejar la elección de 3500 pesetas con el juego indica que el valor 3500 es la mitad del valor 10.000.

Encontremos ahora las cantidades que se igualan con la probabilidad del 50 por ciento de ganar 20.000, 50.000, 100.000 y 200.000 pesetas. Si lo intenta de modo experimental, encontrará normalmente que la cantidad del premio en metálico crecerá en tanto crezca el tamaño de la cantidad apostada. Y observará también que el premio se mantiene aproximadamente proporcional a dicha cantidad, de tal modo que la persona que empareje el premio seguro de 3500 pesetas con el 50 por ciento de 10.000, emparejará 30.000 o 35.000 con la misma probabilidad de ganar 100.000 pesetas. Así, puesto que la cantidad de la apuesta se ha incrementado multiplicando por 10, el premio ha crecido casi igual que si estuviese multiplicado por el mismo factor.

Puede probarse matemáticamente que la función de valor de un individuo que mantiene los premios exactamente proporcionales a las apuestas puede tomar la forma de una función exponencial. En tal función, el valor de una cantidad de dinero equivale a dicha cantidad elevada a cierta potencia. El exponente de la función se puede estimar a través de las elecciones de los sujetos. Si el premio que se iguala a la probabilidad del 50 por ciento de ganar una cierta cantidad es siempre el 35 por ciento de aquella cantidad, el exponente de la función de valor para las ganancias es aproximadamente 2/3. En realidad, la proporcionalidad de los premios y las apuestas es sólo aproximada. En general, el premio aumenta más lentamente que la apuesta. Sin embargo, dado que esta variación en la proporción es pequeña, la función exponencial ofrece una descripción adecuada de los valores de las ganancias.

Puede aplicarse un análisis similar a las pérdidas. Una probabilidad dada de perder 10.000 pesetas o no perder nada se iguala a menudo con una pérdida segura de 4000 pesetas, y una probabilidad dada de perder 100.000 pesetas o nada, con una pérdida segura de 40.000. Si la pérdida segura que se empareja con la probabilidad del 50 por ciento de perder cierta cantidad es siempre el 40 por ciento de esa cantidad, el exponente de la función de valor será aproximadamente 3/4, para el caso de las elecciones entre pérdidas.

Para estudiar conjuntamente las funciones de valor de ganancias y pérdidas, es necesario determinar las perspectivas de riesgo asociadas con los resultados positivos y negativos. Considere un juego en el que le ofrecen la probabilidad del 50 por ciento de perder 10.000 pesetas y una probabilidad del 50 por ciento de ganar un premio en metálico. ¿Cuál será el premio menor por el que acepte usted la apuesta? Es ya un tópico que el placer de ganar una suma de dinero es mucho menos intenso que el miedo de perder esa misma suma. De acuerdo con ello, la gente



FUNCION DE VALOR CONVEXA, que explica las preferencias que tienen que ver con decisiones entre pérdidas. La función asigna un valor negativo a cada pérdida monetaria. Las pérdidas se indican en el eje horizontal, hacia la izquierda del origen; los valores vienen dados en el eje de ordenadas hacia abajo. Al status quo se le asigna el valor cero. La función convexa se hace progresivamente plana para las grandes pérdidas. Son frecuentes las preferencias favorables al riesgo cuando la gente elige entre posibles pérdidas. En una elección favorable al riesgo se prefiere un juego a una pérdida segura que posee una mayor expectativa monetaria. La función de valor predice que un juego que ofrezca la probabilidad de un 85 por ciento de perder 10.000 pesetas y la probabilidad de un 15 por ciento de no perder nada, suele preferirse a una pérdida segura de 8000 pesetas. La expectativa monetaria del juego (−8500) es menor que la pérdida segura (−8000). Sin embargo, el valor del juego (0,85 por el valor de 10.000 pesetas) es mayor (tiene menor magnitud negativa) que el de la pérdida segura. Puesto que el juego tiene mayor valor, se hará preferible. La función de valor convexa, por ello, ayuda a explicar las preferencias favorables al riesgo.

acepta que un juego está equilibrado sólo cuando la posible ganancia es sustancialmente mayor que la posible pérdida. Puede afirmarse, por ejemplo, que la probabilidad de un 50 por ciento de perder 10.000 pesetas es desechable hasta que no se combina con una probabilidad igual de ganar 20.000 o más.

Observando las elecciones que efectúa la gente entre ganancias y pérdidas hemos especificado algunos aspectos de la función de valor. La proporcionalidad aproximada entre premios y apuestas se refleja en la relación exponencial entre las ganancias y sus valores. La aversión al riesgo o la búsqueda del riesgo quedaban descritas por los segmentos cóncavos y convexos de la función de valor. La asimetría entre la ganancia y la pérdida viene expresada por la mayor inclinación de la función de valor para las pérdidas. Importa subrayar que la función de valor es un resumen de un patrón general de elecciones, no una ley universal.

Naturalmente, los sujetos difieren en sus actitudes hacia el riesgo y hacia el dinero, y una única función de valor no puede describir las preferencias de todos los individuos. Las elecciones pueden variar sustancialmente también según el método que se utilice para obtener las preferencias. Por ejemplo, la posibilidad de aceptar una perspectiva arriesgada depende de si el sujeto ha de igualar o emparejar una ganancia con una pérdida dada, una pérdida con una ganancia dada, o ha de variar las probabilidades de riesgo cuando las cantidades de las apuestas son fijas. Además, se rompe la proporcionalidad aproximada de los premios y las apuestas cuando sobrepasamos los valores moderados y entramos en enriquecimientos sustanciales o empobrecimientos relativos. Para ganancias extremadamente grandes, la función de valor se hace prácticamente plana, al no poder los individuos distinguir una enorme ganancia de otra. En la situación negativa, por el contrario, la función suele ser muy inclinada en la proximidad de pérdidas monetarias que forzarían un cambio sustancial de las condiciones o del nivel de vida de las personas.

Al discutir este punto hemos supuesto que la contribución de cada resultado al valor de una perspectiva arriesgada viene ponderada por su probabilidad asociada. Pero hay resultados que evidencian que este supuesto suele violarse, y que los pesos de decisión que multiplican a los valores de los resultados no coinciden con las probabilidades asociadas a estos resultados.

Podemos investigar algunas propiedades de los pesos de decisión mediante el siguiente experimento informal. Imagine que puede mejorar sus elecciones para ganar un premio muy deseable. ¿Pagaría usted tanto para elevar su probabilidad de ganar de un 30 a un 40 por ciento, como para elevarla del 90 por ciento a la certeza total? Generalmente, se está de acuerdo en que la primera oferta atrae menos que la segunda. También se suele estar de acuerdo en que pasar de la imposibilidad a la probabilidad del 10 por ciento es más significativo que el incremento entre el 30 y el 40 por ciento. Así, la diferencia entre certeza y posibilidad, y la diferencia entre posibilidad e imposibilidad, arrojan diferencias mayores que las comparables en los rangos intermedios de probabilidad.

Las investigaciones efectuadas sobre elecciones de riesgo tienden a confirmar esas hipótesis. En relación con la certeza, las probabilidades bajas suelen sobrevalorarse e, infravalorarse, las probabilidades intermedias y altas. Al acontecimiento imposible se le asigna, naturalmente, un peso de cero; y al acontecimiento seguro se le asigna un peso de uno. La sobreponderación de las probabilidades pequeñas puede dar origen a elecciones favorables al riesgo en el terreno positivo (ganancias) y a aversión al riesgo en el terreno negativo (pérdidas), al contrario de lo que ocurría en el patrón de preferencias prevaleciente descrito antes.

El efecto de inflado de las probabilidades pequeñas contribuye al atractivo de los décimos de lotería y los seguros contra accidentes, que se ven en relación con acontecimientos muy significativos, pero relativamente improbables. De otra parte, la infravaloración de las probabilidades intermedias y altas reduce el atractivo de las ganancias posibles respecto de las ciertas. Por la misma razón, también se reduce el miedo a pérdidas posibles en relación con las seguras. Hemos despreciado la contribución de los pesos de decisión en la construcción de la función de valor que describimos más arriba. En aquella función, toda aversión o búsqueda del riesgo se supuso que constituían el resultado de la curvatura de dicha función de valor. Si la probabilidad del 0,5 que se aplicó en la construcción de la función se reemplazase por un peso de decisión ligeramente menor, la función de valor sería menos curva, aunque mantendría la forma general en *S*.

Hasta aquí nos hemos referido a las reglas que gobiernan la evaluación de opciones arriesgadas. Otro importante aspecto de la psicología de las preferencias es cómo la gente define las consecuencias de sus elecciones. La misma decisión puede estructurarse de muy diferentes maneras: diferentes estructuras pueden conducir a diferentes decisiones. Consideremos, por ejemplo, el siguiente problema:

Imagine que, además de lo que posea, se le dan 20.000 pesetas. Se le pide ahora que elija entre (*A*) una ganancia segura de 5000 pesetas y (*B*) una probabilidad del 25 por ciento de ganar 20.000 pesetas y una probabilidad del 75 por ciento de no ganar nada.

La mayoría de las personas a las que se ofrecen estas opciones efectúa una elección adversa al riesgo, prefiriendo la ganancia segura (*A*) sobre el juego (*B*), que tiene el mismo valor esperado. Veamos este otro problema:

Imagine que, además de lo que usted posea, se le ha dado un regalo en metálico de 40.000 pesetas. Se le pide ahora que elija entre (*C*) una pérdida segura de 15.000 pesetas y (*D*) una probabilidad del 75 por ciento de perder 20.000 pesetas y una probabilidad del 25 por ciento de no perder nada.

La mayoría de las personas manifestarían en este caso una preferencia, favorable al riesgo, por el juego (*D*) sobre la pérdida segura (*C*). La opción presentada en los dos problemas es idéntica, sin embargo, en términos objetivos. No hay razón válida para preferir el juego en una versión y el resultado seguro en otra. El elegir la ganancia segura en el primer problema supone una ganancia total de 25.000 pesetas (20.000 de principio más 5000 de la ganancia segura). Elegir la pérdida segura en la segunda versión del problema arroja el mismo resultado, al restar 15.000 pesetas de 40.000. La elección del juego en cualquiera de los dos casos ofrece un 75 por ciento de probabilidades de ganar 20.000 y un 25 por ciento de ganar 40.000.

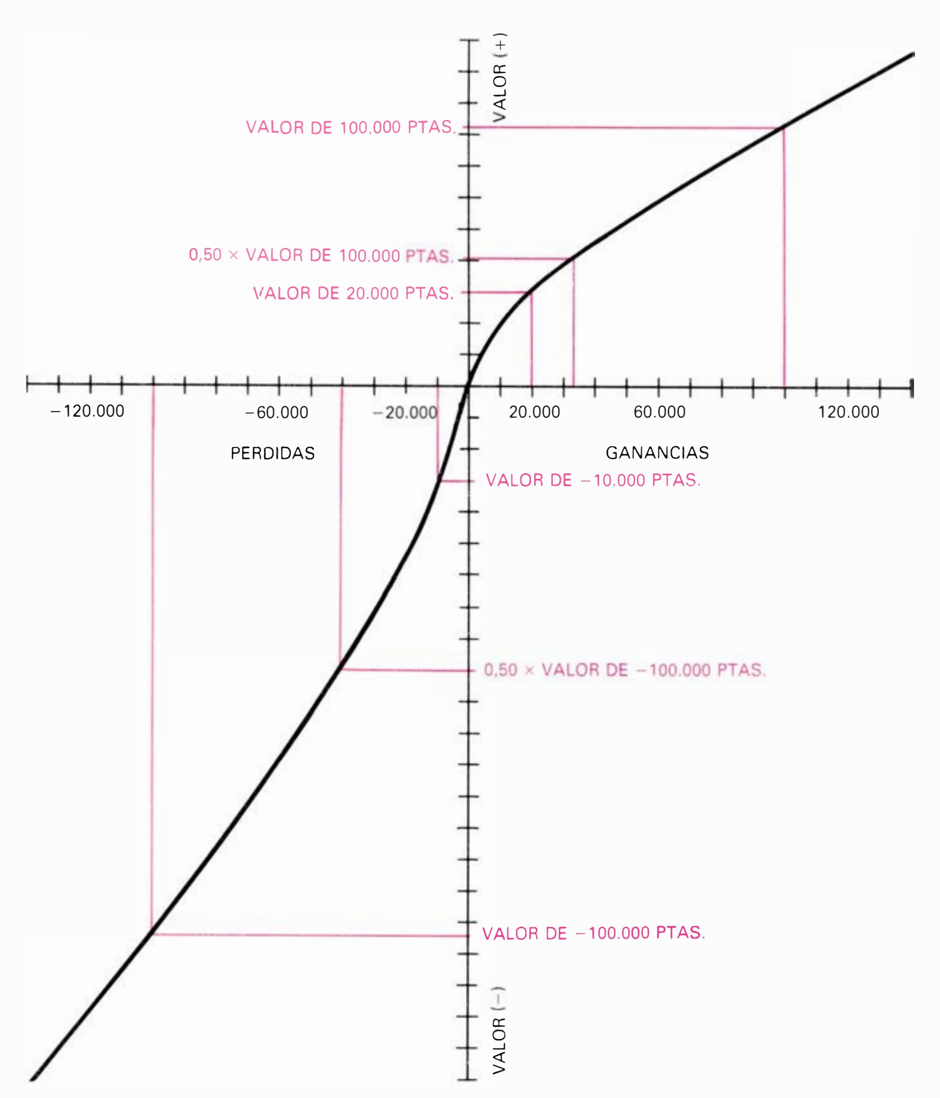
Si los sujetos tuviesen una visión de conjunto de las consecuencias de sus elecciones, tal y como suponen las teorías de la decisión racional, podrían combinar los abonos iniciales con las opciones presentadas y evaluar el resultado compuesto. En lugar de ello, ignoran el abono y evalúan el primer problema como una elección entre ganancias, y el segundo como una entre pérdidas. Alterando la descripción de los resultados se produce la inversión de las preferencias. Llamaremos a esas inversiones efectos de marco.

Los efectos de marco aparecen al evaluar las mismas alternativas objetivas en relación con diferentes puntos

de referencia. Preguntamos a un gran número de médicos para que considerasen el siguiente problema:

Imagine que nos estamos preparando para la llegada de una extraña epidemia que se espera provoque la muerte de 600 personas. Se han propuesto dos programas alternativos para combatir la enfermedad. Supongamos que las estimaciones científicas exactas de las consecuencias de los programas son las siguientes: Si se adopta el programa *A*, se salvarán seguro 200 personas. Si se adopta el programa *B*, hay una probabilidad de 1/3 de que se salven las 600 personas y 2/3, de que no pueda salvarse nadie. ¿Cuál de los dos programas apoyaría usted?

La respuesta de la mayoría a este problema es una preferencia, adversa al riesgo, por el programa *A* sobre el programa *B*.



FUNCION DE VALOR CON FORMA DE *S*. Combina un segmento cóncavo que representa los valores de las ganancias y un segmento convexo representativo de los valores entre las pérdidas. El segmento cóncavo, a la derecha del origen, refleja la aversión al riesgo en las elecciones entre ganancias; el segmento convexo, a la izquierda, refleja elecciones favorables al riesgo entre pérdidas. El segmento de las ganancias ilustra las preferencias en las que una probabilidad del 50 por ciento de ganar una cierta cantidad es tan aceptable como una ganancia segura del 35 por ciento de dicha cantidad. Por ejemplo, una probabilidad del 50 por ciento de ganar 100.000 pesetas es tan aceptable como una ganancia segura de 35.000. El valor de 35.000 es por ello la mitad del valor de 100.000 pesetas. Tal patrón de preferencias puede venir expresado por una función exponencial con un exponente de 2/3 (en una función exponencial el valor de una ganancia es igual a la magnitud de esa ganancia elevada a un exponente fijo). El segmento de las pérdidas representa preferencias en las que una probabilidad del 50 por ciento de perder una cantidad dada es tan aceptable como una pérdida segura del 40 por ciento de tal cantidad: una probabilidad del 50 por ciento de perder 100.000 pesetas es tan aceptable como una pérdida de alrededor de 40.000 pesetas. El valor negativo de −40.000 pesetas es, por ello, la mitad del valor negativo de −100.000 pesetas. Las preferencias relacionadas con las pérdidas pueden expresarse mediante una función exponencial con un exponente de 3/4. La función en forma de *S* ilustra la observación, comúnmente aceptada, de que una pérdida tiene mayor efecto subjetivo que una ganancia equivalente. Puesto que la pendiente de la función para pérdidas es más inclinada que la de las ganancias, el valor negativo de una pérdida de 10.000 pesetas es igual al valor positivo de una ganancia de 20.000. La función, por ello, representa la preferencia de una hipotética persona que hallase aceptable la opción entre ganar 20.000 pesetas o perder 10.000; este es el patrón típico. La función aquí expuesta resume algunos aspectos normales de las preferencias observadas en pruebas con muchas personas; las funciones de valor individuales pueden variar enormemente.

A otros sujetos se les presentó el mismo problema, pero con una formulación diferente de los programas. Si se adoptaba el programa *C* morirían 400 personas. Si se adoptaba el *D*, existiría la probabilidad de 1/3 de que no muriese nadie y una probabilidad de 2/3 de que muriesen las 600 personas.

La elección de la mayoría en este problema es favorable al riesgo: la muerte segura de 400 personas es menos aceptable que una probabilidad de 2/3 de que mueran 600 personas.

Es fácil ver que las dos versiones del problema describen idénticos resultados. La única diferencia estriba en que, en la primera versión, la muerte de 600 personas es el punto de referencia normal, y los resultados se evalúan como ganancias (vidas salvadas), mientras que en la segunda versión el punto de referencia normal es que no haya muertes, y los programas se evalúan en función de vidas perdidas. Puesto que la función de valor tiene forma en *S*, y la certeza se sobrevalora, a través de los dos marcos se producen diferentes preferencias.

Hasta aquí sólo nos hemos referido a decisiones que se enfrentan con una única dimensión. Por el contrario, existen muchas decisiones que se relacionan con la necesidad de efectuar transacciones, en las que los posibles resultados suponen cambios concomitantes en diversas dimensiones de valor. El ejemplo básico de transacción es la adquisición de bienes, tras la cual uno posee más bienes, pero menos dinero que antes. Podemos evaluar las transacciones de acuerdo con el balance de costos y beneficios en una especie de contabilidad mental.

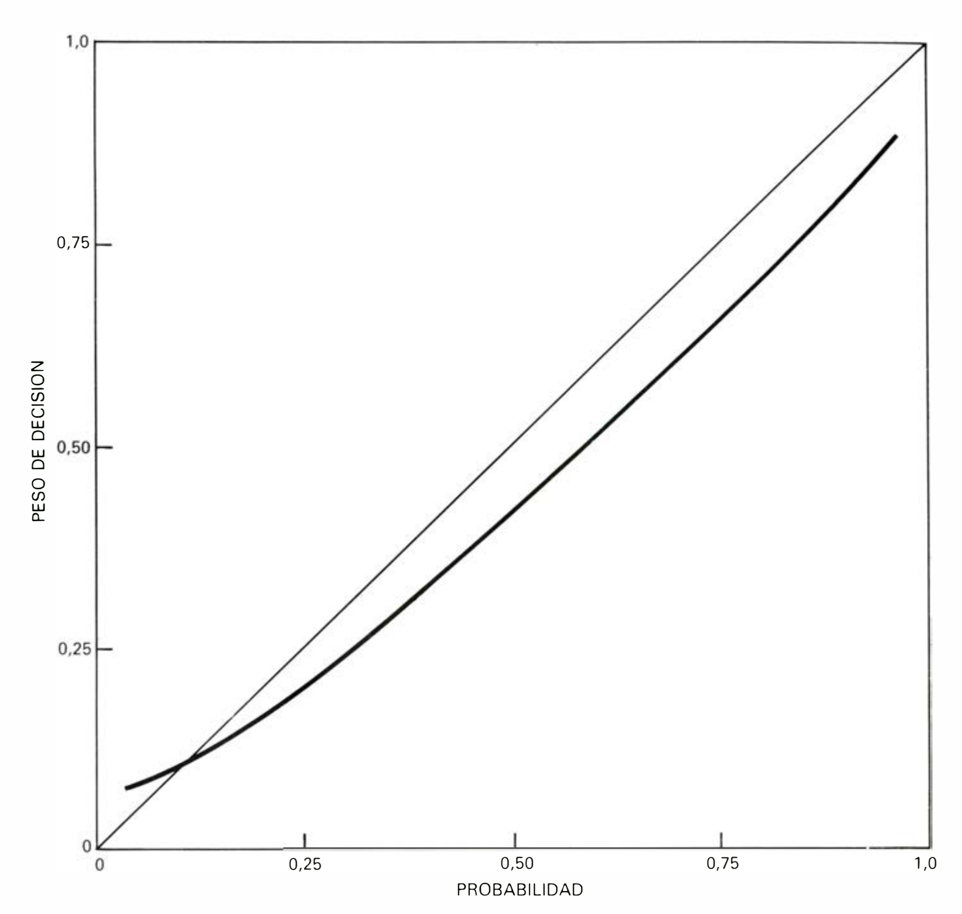
El marco de una transacción puede cambiar de atractivo mediante el control sobre los costos y beneficios que se asignan a cada partida, como ocurría en nuestro ejemplo inicial de los billetes de teatro. Ir al teatro se considera normalmente como una transacción, en la que el precio de la localidad se canjea por la experiencia de ver la obra. Si se ha perdido la entrada, comprar una nueva, efectivamente, dobla el costo de la representación. Por el contrario, la pérdida de dinero no se considera como débito en la partida de la obra. La pérdida de efectivo afecta a la decisión de comprar una nueva localidad en la medida en que una pequeña reducción de riqueza disminuye la tendencia a hacer nuevas adquisiciones opcionales. Cuando presentábamos a una muestra de estudiantes una versión de este problema, la mayoría de aquellos a los que se les decía que imaginasen que habían

perdido el dinero afirmaban que comprarían las entradas, mientras que la mayoría de aquellos a los que se dijo que imaginasen que habían perdido las localidades dijeron que ellos no las comprarían de nuevo.

La contabilidad mental está dirigida por la tendencia a agrupar costos y beneficios alrededor de un objeto, como ilustra el siguiente ejemplo:

Imagine que va a comprar una chaqueta por 12.500 pesetas y una calculadora por 1500. El vendedor de la calculadora le informa de que el modelo que usted quiere comprar está en oferta en otra sucursal de la misma empresa, a 1000 pesetas, pero a 20 minutos en coche del lugar en que se encuentra ahora. ¿Haría usted el viaje a la otra tienda?

La mayoría de los sujetos que contestaron esta pregunta dijeron que ellos viajarían a la otra tienda. Otro grupo respondió a una versión similar en la que se fijó el coste de la chaqueta en 1500 pesetas y el de la calculadora disminuía de 12.500 pesetas en la tienda original a 12.000 en la otra tienda. La mayoría afirmó que no haría el viaje. La compra total y las consecuencias eran idénticas en ambas versiones: el sujeto debía decidir si viajaría 20 minutos para ahorrar 500 pesetas. Las elecciones contrarias de las dos versiones del problema parecen indicar que los sujetos evaluaron el ahorro de 500 pesetas en relación con el precio de la calculadora. En términos relativos, una reducción de 1500 a 1000 pesetas es más impresionante que una reducción de 12.500 pesetas a 12.000.



PESOS DE DECISION: expresan la evaluación subjetiva de las probabilidades. El análisis clásico de las decisiones supone que los pesos de decisión (*trazo grueso*) coinciden con las probabilidades (*trazo delgado*). Por el contrario, investigaciones recientes parecen indicar que los pesos de decisión no coinciden con las probabilidades asociadas a los resultados. Dado que las probabilidades se usan para obtener los valores de los juegos, la respuesta subjetiva a las probabilidades complica la labor de formular una función de valor realista. En particular, la diferencia entre la certeza y la posibilidad y la diferencia entre la posibilidad y la imposibilidad se ponderan en mayor medida que diferencias comparables en los rangos intermedios de probabilidad. Las propiedades principales de los pesos de decisión quedan ilustradas en la curva aquí propuesta. A la imposibilidad se le asigna un peso de cero; a la certidumbre, un peso de uno. Las probabilidades pequeñas se sobrevaloran en relación con la imposibilidad; las propiedades moderadas o grandes se infravaloran en relación con la certeza. La sobrevaloración de las probabilidades pequeñas contribuye al atractivo de los billetes de lotería y las pólizas de seguro, realzando acontecimientos poco probables.

Los efectos de marco en la conducta de los consumidores pueden ser particularmente acusados en situaciones con una única dimensión de costo (usualmente dinero) y varias dimensiones de beneficio. Un magnífico radiocassette es un accesorio muy estimado en la compra de un coche nuevo. Su costo, en cambio, se considera normalmente como un pequeño incremento en el precio del coche. La compra se hace más fácilmente juzgando el valor del cassette independientemente, y su costo como un incremento. Muchos compradores de casas tienen experiencias similares. Los muebles se compran frecuentemente con pocas dudas al mismo tiempo que la casa. Las compras pospuestas, quizá porque no haya existencias de los elementos deseados en el momento de la compra, a menudo parecen costosas y superfluas cuando se contemplan de un modo aislado: su costo resulta demasiado importante por sí mismo. El atractivo de una acción puede, así, cambiar, si su costo o beneficio se carga a una cuenta mayor.

Si reconocemos la influencia de los marcos de referencia con los que se comparan los resultados sobre las decisiones, ¿qué determina a esos marcos de referencia? Es un omnipresente fenómeno psicológico que un punto de referencia depende de las impresiones, juicios y respuestas de los sujetos. El mismo cubo de agua tibia puede sentirse como caliente con una mano y como fría con la otra, si ambas manos han sido previamente expuestas a agua a diferentes temperaturas. Un ingreso recibido puede considerarse pródigo o inadecuado dependiendo de si los propios ingresos han decrecido o crecido recientemente. En esos casos, el punto de referencia es el estado al cual uno se ha ido adaptando. En muchos casos, por el contrario, el punto de referencia viene determinado por acontecimientos que son sólo imaginarios. Consideremos el siguiente incidente:

El señor Pérez y el señor Gómez han programado su salida del aeropuerto en diferentes vuelos al mismo tiempo. Viajaron desde la ciudad en el mismo taxi, sufrieron el mismo atasco de circulación y llegaron al aeropuerto treinta minutos después del horario de salida previsto para sus vuelos. Al señor Pérez se le dijo que su vuelo salió en punto. Al señor Gómez se le dijo que su vuelo salió con retraso, y que acababa de irse hacía sólo cinco minutos. ¿A cuál de los dos altera más el incidente?

Casi todos los sujetos a los que se presentó esta situación piensan que el señor Gómez estará más apenado que el señor Pérez, aunque sus condiciones objetivas sean idénticas: ambos han perdido su vuelo. Más aún, ambos esperaban perder su vuelo, de tal modo que el señor Gómez no tiene razón para sentirse más decepcionado que el señor Pérez. Si el señor Gómez es el más desilusionado, ello se debe, presumiblemente, a que su imaginación le per-

mite acercarse más que el señor Pérez a la posibilidad de tomar el vuelo a tiempo. La frustración experimentada en una situación insatisfactoria se incrementa cuando es fácil imaginar una alternativa más deseable. Consideremos otro ejemplo similar:

El número ganador de la lotería fue el 865.304. Tres individuos comparan el billete que poseen con el número ganador: Juan tiene el 361.204, Isabel el 965.304, y Pedro tiene el 865.305. ¿Cuál será su nivel respectivo de pesadumbre?

Hay acuerdo general en que la experiencia será devastadora para Pedro, bastante seria para Isabel y muy suave para Juan. Aquí, otra vez, la ordenación corresponde al nivel de cercanía a ganar el premio en que los individuos consideraron hallarse.

La experiencia de un individuo sobre el placer o la frustración puede, por ello, depender de un acto de imaginación que determina el nivel de referencia con el que se compara la realidad. Es digno de tener en cuenta que el acto de imaginación por el que uno crea realidades alternativas refleja muchas compulsiones reales, u ocasionadas por la realidad en sí. Si no existiesen estas fuerzas, el señor Pérez encontraría tan fácil como el señor Gómez el imaginarse a sí mismo tomando el avión, y Juan podría encontrar tan sencillo como Pedro el imaginarse con el billete ganador. La imaginación, así, parece estar gobernada por ciertas reglas, y tales reglas afectan a nuestra experiencia de la realidad, controlando las alternativas con las que se compara.

El remordimiento es una forma especial de frustración en la que el acontecimiento que uno podría cambiar depende de una acción que se ejecutó u omitió. Una prolongación natural de la hipótesis que hemos venido desarrollando para el análisis de la frustración es que se siente remordimiento si uno puede imaginarse fácilmente efectuando una acción que le podría haber llevado a un resultado más deseable. Esta interpretación explica la íntima relación entre la experiencia del remordimiento y la autodeterminación al elegir: las acciones ejecutadas compulsivamente generan poco remordimiento. La renuncia a violar procedimientos típicos y a actuar de modo original o innovador puede también ser una defensa efectiva contra el remordimiento subsiguiente; es fácil imaginarse ejecutando actos convencionales y más difícil haciendo lo no convencional.

Una hipótesis directamente relacionada con lo anterior es que resulta a menudo más fácil suprimir un acontecimiento de una cadena de hechos que imaginar la inclusión de un acontecimiento en la cadena. Tal diferencia en la capacidad de imaginación puede ayudar a explicar que el remordimiento asociado a las omisiones de actos es frecuentemente menos intenso que el asociado a los errores al actuar:

Julián posee acciones en la Compañía *A*. Durante el pasado año estuvo pensando en cambiarlas por las de la Compañía *B*, pero decidió no hacerlo. Se da cuenta ahora de que podría haber ganado 120.000 pesetas de haberlo hecho. David tenía acciones de la Compañía *B*. Durante el pasado año las cambió por las de la Compañía *A*. Se encuentra ahora con que podría haber ganado 120.000 pesetas si las hubiese mantenido en la Compañía *B*.

Aquí se está generalmente de acuerdo en que David está más apenado que Julián, aunque su situación objetiva sea idéntica (ambos tienen acciones de la Compañía *A*), y los dos llegan a esta situación por decisión propia.

Aparentemente es más fácil por parte de David imaginarse no llevando a cabo la acción (y, por lo tanto, manteniendo el paquete de acciones más ventajoso) de lo que sería para Julián imaginarse habiendo actuado. Más aún, podría esperarse que ambos hubiesen anticipado la posibilidad del remordimiento, y actuado en consecuencia. En general, es probable que la anticipación del remordimiento favorezca la inacción sobre la acción, y la conducta de rutina sobre la innovadora.

Hemos considerado una gran cantidad de ejemplos en los que una decisión, una preferencia o una reacción emocional eran controladas por factores que pueden parecer irrelevantes. Alguno de esos ejemplos, que fueron utilizados por Herbert A. Simon bajo la denominación de "racionalidad limitada", manifiestan obstáculos para el logro de la elección racional. La dificultad que tienen las personas para mantener un punto de vista integrado de las consecuencias de sus acciones y elecciones, y su susceptibilidad a las variaciones de los marcos de referencia, constituyen ejemplos de tales impedimentos. El estudio descriptivo de las preferencias presenta también retos a la teoría de la elección racional, porque ésta, con frecuencia, no aclara bien si los efectos de los pesos de decisión, puntos de referencia, marcos y remordimientos han de considerarse como errores o sesgos, o más bien han de aceptarse como elementos válidos de la experiencia.

Temas metamágicos

Un artículo auto-alusivo referente al artículo sobre auto-alusión del pasado mes de marzo

Douglas R. Hofstadter

Por ser éste el número de marzo, he pensado en retomar el hilo de mi artículo de hace un año, que dediqué entonces a las construcciones auto-alusivas, y tratar nuevamente aquí ese tema. No obstante, quiero aprovechar ya el párrafo inicial para prevenir a los lectores que no gusten de la sibireferencia de que tal vez estarán deseando dejar de leer antes de terminar este párrafo, quiero decir, esta frase, o sea, esta cláusula, o incluso este sustantivo –en una palabra, esto.

Desde la publicación de aquel artículo me ha llegado una crecida remesa de correspondencia auto-alusiva. Astutamente, Tony Durham se dejaba caer diciendo "... que con el volumen de respuestas previsible, no creo que sea usted en persona quien lea esto". La carta de John C. Waugh gañía: "¡Socorro! ¡Estoy sepultada bajo una avalancha de cartas de lectores!". Su nota fue rescatada y ahora reposa en una pila mucho más reducida. Ofreceré aquí algunas de mis favoritas de esa colección.

Mas no quiero dejar el tema del correo sin mencionar que el matasellos de la postal que Ivan Vince me dirigió desde Inglaterra declaraba crípticamente: "Be properly addressed" (Indique las señas correctamente). ¿Se trataba de una orden que la Oficinal Postal dirigía a la propia tarjeta? De ser así, las tarjetas inglesas han de ser bastante más inteligentes que las demás; por acá no hemos dado todavía con tarjetas que sepan leer y, no digamos, corregir las señas a que se las destina. (¡Una de las tarjetas que me llegó iba dirigida a la revista *Omni*!)

Quedé muy halagado por un par de elogios auto-invalidantes: "Sencillamente, no puedo expresarle cuánto me gustó su primer artículo", me escribía Richard Ruttan. Y John Collins declaraba: "Esta nota no comunica con qué delicia leí la sección de marzo". También me resultó gratificador saber que mi fama se había extendido nada menos que hasta los lavabos de caballeros de la Facultad de Filosofía de la Universidad Tufts, en una de cuyas paredes Dan Dennett halló garrapateada: "Esta frase es una pintada.-Douglas R. Hofstadter".

Uno de los pasatiempos favoritos de mis corresponsables ha sido la búsqueda de preguntas interesantes que se den respuesta a sí mismas. Tan sólo unas cuantas han conseguido "salirse del sistema", lo que para mí es lo mismo que ser verdaderamente nuevas. Según parece, no es fácil alcanzar éxitos en esta microscópica forma artística. John Flagg hizo notar con cinismo (he parafraseado ligeramente su original): "Formule una pregunta que se responda a sí misma, y tendremos una respuesta auto-interrogante". Una de mis favoritas la dio Henry Taves: "Recuerdo con cariño un examen de historia que me pusieron cuando estaba en un internado, y que contenía la pregunta siguiente: IV. Redacte una pregunta adecuada para un examen final de esta materia y contéstela. La respuesta que di fue, sencillamente, copiar dos veces esa frase". Quedé encantado al leer esto, pero luego, reflexionando, empecé a pensar que hay en ella algo que no va bien del todo. ¿Qué le parece a usted?

Richard Showstack contribuyó con dos burlonas frases auto-respondidas: "¿Qué pregunta no verbo?" y "¿En qué pregunta se menciona la palabra 'paraguas' sin que haya motivo?". Jim Shiley aportó una aguda frase que yo he modificado ligeramente: "¿Es ésta una pregunta retórica, o es ésta una pregunta retórica?". También suya es ésta: "Tome una cuartilla en blanco y escriba en ella: '¿Hasta dónde ocupará esta pregunta en la cuartilla?'. Ahora, si alguno de sus amigos políglotas hace observar que en idioma ural-altaico esa misma secuencia de grafemas significa 'unos quince centímetros', creo tener derecho a una suscripción gratuita a *Investigación y Ciencia*. En caso contrario, si la inscripción de una pregunta cuenta tanto como pregunta cuanto como unidad de medida, creo que merezco al menos un premio de consolación. Pero me parece que he forzado un poco las reglas".

Mis propias soluciones al problema de la pregunta auto-respondida resultan no ser tanto preguntas que se auto-*responden* cuanto preguntas que se auto-*provocan*. Así ocurre en la siguiente: "¿Por qué me sales tú ahora con *ésas*?". Salta a la vida que cuando se hace una pregunta así es bastante probable que nuestro interlocutor *nos* salga con una pregunta idéntica, muestra de su perplejidad. De acuerdo, sí, pero ¿qué provoca *éste* interrogante?

Por intermedio de Philip Cohen supe de la siguiente anécdota acerca de una pregunta auto-respondida debida a Damon Knight: "Un viejo amigo, Terry Carr, nos envió un acertijo en una postal, y más tarde, en otra, nos dio la solución. Entonces nos envió otro más, que preguntaba: '¿Cómo dejar intrigado a un pardillo?' De ésta no dio solución. Estuvimos esperando un par de semanas antes de caer en la cuenta."

Varias de las auténticas obras maestras que he recibido durante el año pertenecen a la categoría que llamo "auto-documentada". Vemos un ejemplo sencillo en "Esta frase tiene doce palabras, veintiseis sílabas y sesenta y cuatro letras", original de Jonathan Post. He aquí un airoso giro, de la pluma de John Atkins: "'Formada por veintiseis letras' lo está". Las formas auto-documentadas pueden hacerse mucho más intrincadas e introspectivas. Philip Cohen me hizo saber de un ejemplo debido a Howard Bergerson, maestro en juegos de palabras. Reza así: "En esta frase, la palabra APARECE aparece catorce veces, la palabra CATORCE aparece cuatro veces, la palabra CUATRO aparece repetida, la palabra EN aparece repetida, la palabra ESTA aparece repetida, la palabra FRASE aparece repetida, la palabra LA aparece catorce veces, la palabra OCHO aparece repetida, la palabra PALABRA aparece catorce veces, la palabra REPETIDA aparece ocho veces, la palabra SIETE aparece repetida, la palabra VECES aparece siete veces, y la palabra Y aparece repetida."

No está nada mal, pero la medalla de oro de esta categoría está reservada a Lee Sallows, quien presentó el *tour de force* siguiente (y no muy sencillo de traducir): "Only the fool would take trouble to verify that his sentence was composed of ten a's, three b's, four c's, four d's, forty-six e's, sixteen f's, four g's, thirteen h's, fifteen i's, two k's, nine l's, four m's, twenty-five n's, twenty-four o's, five p's, sixteen r's, forty-one s's, thirty-seven t's, ten u's, eight v's, eight w's, four x's, eleven y's, twenty-seven commas, twenty-three apostrophes, seven hyphens and, last but not least, a single!"

Yo (tal vez haciendo el tonto) me tomé la molestia de verificar lo que en la frase se decía. No obstante, antes de lanzarme hice unas cuantas comprobaciones al azar. Y tengo que decir que la primera verificación tuvo un fuerte efecto psicológico: súbitamente, la valoración de credibilidad de la frase creció muchos puntos. Y me llamó la atención, por fantástico (y maravilloso) cómo en ciertas situaciones la verificación de un pequeño tanto por ciento de una teoría puede servir para reforzar poderosamente nuestra confianza en la validez de su totalidad. ¡Tal vez sea éste todo el objetivo de la frase!

El afamado lógico Raphael Robinson envió un travieso rompecabezas, también de la serie "auto-documentada". Se pide del lector que rellene correctamente los espacios en blanco de la siguiente frase: "En esta frase, el número de veces que aparece el 0 es _____, el 1 es _____, el 2 es _____, el 3 es _____, el 4 es _____, el 5 es _____, el 6 es _____, el 7 es _____, el 8 es _____ y el 9 es _____." Cada espacio en blanco debe quedar ocupado con una cifra (de uno o varios dígitos) en notación decimal. Hay exactamente dos soluciones. Los lectores podrían también esforzarse en hallar dos frases que se documenten mutuamente, o bucles de este tipo de tamaño mayor.

Evidentemente, el súmmum de las frases auto-documentadas no podrán alcanzarlo frases que se limiten a inventariar sus componentes; para ello harán falta más bien frases que contengan, además, una regla que explicase *urbi et orbi* cómo volver a ensamblar sus partes componentes para crear la frase completa en breve, una frase partenogenética, auto-reproductora. Disponemos de una frase así: la versión (castellana) de la versión (inglesa) que Willard Van Orman Quine dio del clásico homenaje metamatemático que Kurt Gödel rindió a Epiménides el Cretense:

"produce falsedad al agregarla tras su referencia" produce falsedad al agregarla tras su referencia.

La frase de Quine, en efecto, le dice al lector cómo construir una reproducción de la frase recién leída, y entonces (sólo por completar la cosa) añade que la copia (¡no *ella* misma, quede claro!) afirma una falsedad. Nos recuerda la famosa observación que Epilópsides el Concretense (primo segundo de Epiménides) le hizo a Flora, hermosa y joven mujer cuyo ardiente amor él no podía corresponder (por hallarse comprometido con Fauna, hermana gemela de Flora): "Ten ánimo, querida mía. Voy a proponerte algo que tal vez te alegre. Toma una de las células de mi brazo y de ella haz un cultivo clónico. Al fin tendrás como pareja un ser de aspecto y sentimientos idénticos a los míos. Pero cuidado, no te fíes de él. ¡Tiene propensión a contarles bolas fenomenales a las mujeres bonitas!"

En la década de los cuarenta, John Von Neumann hizo un gran esfuerzo para diseñar una máquina capaz de construir una copia de sí misma a partir de ciertos materiales primarios. Al cabo consiguió un diseño teórico, formado por cientos de miles de componentes. Vista en retrospectiva, y con grado de abstracción considerable, la idea implícita en la máquina autoreproductora de Von Neumann resulta notablemente similar al método de que se vale el ADN para producir copias de sí mismo. Y éste método se halla próximo al empleado por Gödel para construir una proposición auto-alusiva en el seno de un lenguaje matemático, donde, en principio, no parece haber forma de aludir al lenguaje mismo.

Por ello, tenemos el honor de presentar el Gran Trofeo Von Neumann, disputado un decenio sí y otro no, que, pensado para los más ambiciosos de nuestros lectores, pide la creación de frases auto-documentadas, comprensibles y cuya longitud no esté fuera de razón, en la que no sólo se dé la relación de sus elementos componentes (sea a nivel de letras o palabras) sino que explique además cómo deben reunirse y ensamblarse sus partes de forma que la frase quede reconstituida. (Por cierto, fijémonos en que la condición requerida es que la longitud de la frase no esté fuera de razón, lo que es muy diferente de que sea razonablemente larga.) La relación de componentes (las "semillas") debería ser un inventario de palabras o símbolos tipográficos, poco más o menos igual que en las frases de Bergerson o Sallows. Los signos inventariados deberían ser claramente distinguibles del texto que los enlaza y que a ellos se refiere. Podrían, por ejemplo, encerrarse entre comillas, o imprimirse con distintos caracteres, o mencionarse por su nombre. Poco importa el convenio concreto que se adopte con tal de que la distinción sea tajante. El resto de la frase (la "regla de construcción") debería quedar impreso en texto ordinario, pues no ha de ser considerado como materia prima tipográfica, sino como sistema de instrucciones de montaje. Nos encontramos nuevamente aquí la distinción entre uso y mención que ya discutí en mi artículo anterior; de no tenerla en cuenta se cometería un error conceptual grave. (Así ocurre en la frase de Sallows, empañando levemente el oro de su medalla.)

La regla de construcción quizá no aluda al material impreso en forma ordinaria, sino solamente a ciertas partes del inventario. Por consiguiente, ¡no es lícito que la regla de construcción se mencione a sí misma para nada! La regla de construcción tiene que describir explícitamente la estructura. Además, y tal vez sea éste el aspecto de la autoalusión más frecuentemente olvidado, la regla de construcción ha de especificar qué partes habrán de imprimirse normalmente y qué otras deberán entrecomillarse o distinguirse de alguna otra manera. En este aspecto, la frase de Bergerson falla. Aunque tiene el mérito de distinguir tajantemente entre uso y mención, empleando letra mayúscula para los elementos del inventario y letra minúscula para denotar los recuentos y las palabras de empaste y enlace, no presenta en cambio inventarios independientes de los componentes en mayúsculas y en minúsculas. Por el contrario, su frase los contabiliza todos juntos, echando así a perder una distinción fundamental.

En el Gran Trofeo Von Neumann se concederán puntos extra a las soluciones que vengan redactadas en lenguaje natural básico y, también, a las soluciones cuya simiente se mantenga al nivel de letra (como en la frase de Sallows). Aunque es evidente que la frase de Quine contiene una semilla (el grupo de siete palabras entrecomilladas) y una regla de formación (que estipula la agregación de algo a su referencia), no podemos lícitamente dejarla concursar, porque su semilla dista mucho de ser materia prima en bruto.

Y ya que hablamos de ella, hay una buena razón para que la semilla de la frase de Quine sea tan complicada –más aún, para ser idéntica a la regla de construcción salvo por las comillas. Esta razón es fácil de enunciar: es necesario *construir la regla de construcción* a partir de las materias primas, y por ello, cuanto más semejante sea la regla de construcción a la semilla tanto más sencilla resultará su elaboración. Para obtener una frase enteramente nueva todo cuanto se precisa es hacer dos copias de la semilla, llevar a cabo cualesquiera manipulaciones sencillas que conviertan una de las copias en regla de construcción, y después acoplar el otro ejemplar de la semilla a la regla de construcción recién acuñada. Tendremos formada así una frase nueva y completa, lista para salir de la cadena de montaje.

Para aclarar un poco todo esto puede servirnos de ayuda un ligera variante de la frase de Quine. Imaginemos que solamente supiéramos leer con letras minúsculas, y que los caracteres mayúsculos nos fueran desconocidos. Entonces, los textos impresos con ellos nos serían incomprensibles, y carecerían de significado e interés; los impresos con minúsculas podrían en cambio sugerirnos ideas o acciones. Supongamos en este punto que alguien nos proporcionase una tabla de conversión que asociase cada letra mayúscula con su correspondiente minúscula; ello nos permitiría decodificar los textos en mayúsculas. Un buen día nos encontramos con este trozo de texto en mayúsculas, que no sabemos leer directamente:

PRODUCE FALSEDAD AL UTILIZARLO COMO SUJETO DE SU VERSIÓN CON MINÚSCULAS

Una vez descifrado, el texto producirá una frase escrita con minúsculas, o mejor dicho, un fragmento de frase en letra minúscula, un predicado sin sujeto. Sugestivo, ¿verdad?

La idea de manejar paralelamente dos alfabetos, en uno de los cuales los textos son inertes y desprovistos de significado, mientras en el otro son activos y significantes, podrá tal vez parecerle solamente una pequeña variación de la frase de Quine, pero en realidad es muy semejante a un recurso excepcionalmente ingenioso que la naturaleza ha descubierto y explotado en todas las células de los seres vivos. Nuestra semilla –nuestro ADN– es un enorme libro de texto inerte escrito con un alfabeto químico formado por 64 letras "mayúsculas" (los codones). Nuestras reglas de construcción –nuestros enzimas– son mensajes breves y expresivos, texto activo escrito con alfabeto químico diferente formado tan sólo por 20 letras "minúsculas" (los aminoácidos). Hay una tabla (el código genético) de conversión de letras mayúsculas en minúsculas. Como es evidente, algunas minúsculas habrán de ser traducción de más de una mayúscula, pero hay un detalle a tener en cuenta. Resulta también que tres caracteres del alfabeto de mayúsculas no actúan como letras, sino como signos ortográficos, que indican dónde termina un mensaje y empieza otro. Una vez conocida la tabla de conversión, suele suceder que no nos acordemos siquiera de distinguir entre ambos tipos de alfabetos químicos: el inerte, en mayúsculas, compuesto de codones, y el alfabeto de las minúsculas, activo, formado por aminoácidos. Lo esencial es que, armados del código genético, podemos leer el libro del ADN (la semilla) como si fuera una sucesión de mensajes enzimáticos (reglas de construcción) que explican cómo construir un nuevo libro de ADN y un sistema nuevo de mensajes enzimáticos. El paralelismo con la variante de la frase de Quine es perfecto. En ella, el texto inerte, en mayúsculas, se convirtió en texto en minúsculas, que da la regla para lograr un ejemplar de la frase de Quine completa a partir de su semilla.

El ADN y los enzimas celulares actúan como las semillas y las reglas de construcción de la frase de Quine, o la lista de componentes e instrucciones de formación del autómata autorreproductivo de Von Neumann, o también, como las semillas y reglas de formación de programas de ordenador capaces de imprimirse a sí mismos. Resulta asombroso cuán universal es el mecanismo de auto-alusión, y por ello, siempre me choca que las mismas gentes que vituperan de la supuesta estupidez de la auto-alusión estén ellas mismas compuestas por trillones de moléculas auto-referidas.

Scott Kim y yo hemos descubierto un extraño par de frases:

"La frase siguiente es por completo idéntica a ésta, excepto en que las palabras 'siguiente' y 'anterior' están intercambiadas, y también lo están 'excepto' y 'en', lo mismo que las locuciones 'idéntica a' y 'diferente de'."

"La frase anterior es por completo diferente de ésta, en que las palabras 'anterior' y 'siguiente' están intercambiadas, y también lo están 'en' y 'excepto', lo mismo que las locuciones 'diferente de' y 'idéntica a'."

A primera vista estas frases nos traen reminiscencias de la paradoja de Epiménides, como si fuera una versión "dúplex" de aquélla ("La frase siguiente es verdadera"; "La frase anterior es falsa"). No obstante, en la segunda ojeada se ve que cada una dice exactamente lo mismo que la otra. Don Byrd no está en esto de acuerdo conmigo; según él dicen cosas completamente distintas.

No sorprende, por tanto, que varias de las frases presentadas por los lectores exhalen aromas de paradoja. Algunas son variantes de la paradoja de Russell, que habla del barbero que afeita a todos quienes no se afeitan a sí mismos, o del conjunto de todos los conjuntos que no se cuentan a sí mismos como elementos. Por ejemplo, Gerald Hull preparó esta extraña concocción: "Esta frase alude a toda frase que no aluda a sí misma". ¿Será autoalusiva la concocción de Hull, o no? En vena parecida, Michael Gardner citó la dedicatoria de la tesis de un compañero: "Esta tesis está dedicada a todos aquellos que no se dedicaron sus tesis a sí mismos". Charles Brenner me hizo notar que en el libro *Model Theory* de C. C. Chang y H. J. Keisler hay una dedicatoria parecida. El mismo Brenner propuso otra variante de la paradoja de Russell: "Redacte un programa de ordenador que encargue a la máquina imprimir todos los programas que en ningún momento hayan de imprimirse a sí mismos." La cuestión, evidentemente, es si el programa llegará a imprimirse a sí mismo, o no.

Una de las frases más desorientadoras me llegó de Robert Boeninger: "Esta frase posee realmente la propiedad de no tener la propiedad que declara no tener". ¿Lo ha entendido? Bien, el problema es justamente averiguar qué propiedad es ésa de que la frase declara carecer.

El matemático holandés Hans Freudenthal me hizo saber de una paradójica anécdota basada en la auto-alusión:

"Hay un cuento alemán del siglo XVIII, debido a Christian Gellert, llamado 'Der Bauer und sein Sohn' (El labriego y su hijo). Un día, yendo de camino, el hijo cuenta al padre una gran mentira. Su padre, ladinamente, le hace saber que se están acercando al 'puente de los mentirosos', puente que se hunde siempre que un mentiroso lo cruza. Aterrado por el relato de su padre, el muchacho admite haber mentido, y confiesa la verdad.

"Cuando yo [Freudenthal] le conté a un chaval de 10 años este cuento, el chaval quiso saber qué sucedió cuando llegaron al puente. Yo contesté: 'Se hundió al pasar el padre, por mentiroso, pues en realidad no hay puentes de mentirosos' (¿O tal vez sí?)"

C. W. Smith, quien me escribió desde London (Ontario), expuso una situación que recuerda también a la paradoja del mentiroso:

"Durante los años sesenta, plantado solitario en medio de un solar cubierto de hierbajos, había un anuncio roído por la intemperie que decía: '25 dólares de recompensa por toda información que conduzca a la detención y condena de quien se lleve este anuncio'. Valga lo que valiere, hace mucho que el anuncio desapareció. Y también el solar".

A propósito: la paradoja de Epiménides el Cretense no debe confundirse con la paradoja de Nixónides el Cretino (hacia 1974 d.C.): "Esta declaración es inválida". Y hablando de Epiménides, una de las más elegantes variaciones de su famosa paradoja está en la sección

de erratas de un libro hipotético descrito por Beverley Rowe. He aquí su aspecto:

(vi)
Errata
Página (vi): Donde dice "Errata" debe decir "Erratum"

Parientes próximas de las frases genuinamente paradójicas son las que llamo frases "neuróticas" y frases "cuerdas". Las frases cuerdas son las que, por así decirlo, llevan a la práctica lo que declaran, mientras que las neuróticas son frases que dicen una cosa pero practican lo contrario. Alan Auerbach nos ha dado un buen ejemplo de cada categoría. Su frase cuerda es "¡Breve!". Su frase neurótica es "El buen estilo literario –lo hemos oído un millón de veces– rehúye la exageración". He aquí una frase cuerda de Brad Shelton: "Hace ochenta y siete palabras esta frase no había comenzado todavía". Una de las frases más originales es de Carl Bender: "La siguiente frase está escrita en Tailandia, va".

Fijémonos en una frase emparentada con ella, enviada por David Stork: "No es preciso decir que...". ¿A qué categoría pertenece? Tal vez sea una frase psicótica. Pete Maclean ha enviado una que puede dejarnos perplejos: "Si 'verdadera' y 'falsa' permutasen mutuamente sus significados, esta frase no sería falsa". ¡Aún no tengo claro qué significa! Dan Krimm me escribió entre otras: "He oído decir que esta frase es un rumor". Linda Simonetti contribuyó con la siguiente: "Que no es una frase completa, sino tan sólo una oración subordinada". Douglas Wolfe nos ofrece la "regla neurótica de perceptiva literaria", a saber: "Nunca emplee el modo imperativo; tampoco es nunca correcto construir una frase mezclando en ella varios modos del verbo". David Moser me recordó un eslogan que el *National Lampoon* utilizó un tiempo "Es tan divertido que no necesitamos eslogan". Perry Weddle me contó: "Estoy tratando de lograr que mi loro diga 'No comprendo nada de cuanto digo'. Cuando lo digo *yo* la frase es perversamente auto-alusiva, pero, ¿y cuándo lo dice el loro?". Stephen Coombs hizo notar que "una frase puede auto-aludirse en el verbo". Mi madre, Nancy Hofstadter, ha oído al Secretario de Estado Haig calificar un mensaje de advertencia a los rusos de "calculada ambigüedad que sin duda será perfectamente comprendida". ¡Sí, mi general!

Jim Propp ha enviado una sucesión de frases que van deslizándose desde la cordura neurótica hacia la neurosis cuerda:

"Esta frase cada tercera, pero todavía comprensible."

"Esto más fácil comprender menos hubiera omitido."

"Esto imposible salvo contexto."

"Idea intento 4′33″."

" "

La penúltima frase se refiere a una famosa pieza musical de John Cage (transcripción para piano) que consiste en cuatro minutos y 33 segundos de silencio. La última frase pudiera muy bien ser una cita de *The Wit and Wisdom of Spiro T. Agnew* (Ingenio y sabiduría de Spiro T. Agnew), pero la verdad es que es demasiado breve para asegurarlo con certeza. Propp envía también la cuerda cita siguiente, tomado de *Intelligence in Ape and Man*, de David Premack: "Por 'productividad' de un lenguaje yo entiendo la capacidad del lenguaje para crear nuevos términos cuyo significado se explica gracias a los antiguos".

El filósofo Howard DeLong contribuyó con lo que pudiera llamarse un "silogismo neurótico", que reza así: "Todo silogismo inválido infringe al menos una regla. Este silogismo infringe al menos una regla. Por consiguiente, este silogismo es un silogismo inválido".

Diversos lectores me han hecho saber de frases felices que han tenido éxito últimamente. D. A. Treissmann, por ejemplo, me recordaba que "la nostalgia ya no es lo que era". Henry Taves mencionaba las graciosas camisetas adornadas con frases como "Mi gente se piró a Florida y todo lo que me trajeron fue esta piojosa camiseta". Y John Fletcher me contó de un episodio del programa de televisión "Laugh-In", hace de ello ya algunos años, en el que Joanne Worley cantaba "No soy más que una chica que ni siquiera sabe decir 'n...', 'n...', 'n...'". John Healey me escribió: "Antes yo solía pensar que era un indeciso, pero ahora ya no estoy tan seguro de ello".

Yo mismo puedo aportar algo a esta colección. He aquí una frase neurótica: "En esta frase, las tres últimas palabras 'han quedado fuera'". ¿De veras es neurótica? ¡Estas cosas me confunden! Sea como fuere, esta frase es totalmente cuerda: "Esta frase ofrece a su(s) lector(a, es, as) diversas alternativas/opciones que puede(n) libremente aceptar y/o rechazar". Tenemos después la inevitable "Esta frase es neurótica". Lo malo es que si realmente *es* neurótica, entonces practica lo que predica y por tanto es cuerda y *no puede* ser neurótica; aunque en tal caso, si *no* es neurótica entonces es lo contrario de lo que dice ser, y por tanto *sí* tiene que ser neurótica. ¡No es milagro que esté neurótica!

Y hablando de frases neuróticas, ¿qué decir de las frases con crisis de identidad? Estas son para mí las más interesantes de todas. Un ejemplo típico es la pregunta (henchida de vaga aprensión) que ha preparado David Krimm: "Si yo afirmase alguna otra cosa; ¿sería yo misma todavía?" Me pareció posible perfeccionarla un poco, así que la modifiqué ligeramente: "Si yo dijese alguna otra cosa, ¿sería todavía yo quien lo dice?" No quedando satisfecho, redacté una nueva variación: "En otro mundo, ¿hubiera yo podido ser una frase acerca de Humphrey Bogart?" Cuando me detuve a reflexionar en lo escrito caí en la cuenta de que al reelaborar la frase de Krimm había estado trasteando su personalidad justamente del modo que expresa temer. La pregunta subsistía. Allá en lo más profundo, ¿eran estas frases realmente una misma frase? Mi último experimento en este sentido fue "En otro mundo, ¿podría esta frase haber sido la frase de Dan Krimm?"

Como es obvio, algunos lectores estuvieron explorando líneas paralelas a ésta. John Atkins demandaba: "¿Puede alguien explicar por qué esta revista seguiría siendo la misma sin esta pregunta, y sin embargo, esta pregunta no sería la misma sin esta palabra?" Y Loul McIntosh, que trabaja en un centro de rehabilitación para esquizofrénicos, tenía una pregunta que relaciona la identidad personal con la auto-alusión: "Si yo fuera usted, ¿quién estaría leyendo esta frase?" Al pie de su carta añadía ella: (N. B.: Esto es lo que me pasa por trabajar con esquizofrénicos"). Lo cual me lleva a Peter M. Brigham, médico, quien en su trabajo se tropezó con un caso de esquizofrenia literaria: "Evidentemente, acaba usted de empezar a leer la frase que acaba de terminar de leer".

No hace mucho oí un programa radiofónico sobre animales domésticos en el que el locutor comentaba: "Si el guión de este programa hubiera sido escrito por un perro, seguramente diría que las personas son inferiores a los perros, porque no pueden mover el rabo." Lo que me dio pie para pensar cómo hubiera podido ser esta sección si la hubiera escrito un perro. No puedo asegurarlo, pero me da la impresión de que estaría dedicada a la persecución de ardillas. Y tal vez contuviera un pá-

rrafo que especulase cómo sería esta sección de haber sido escrita por una ardilla.

Creo que entre todas las frases recibidas, mi favorita es la enviada por Harold Cooper, frase que fue inspirada por mi pregunta auto-alusiva y contrafáctica del año pasado: "¿Qué aspecto tendría esta frase si π tuviera el valor 3?" He aquí su respuesta:

Si π fuese igual a 3, esta oración tendría un aspecto así

Las tres oes hexagonales denotan el hecho de que la relación de la "circunferencia" del hexágono a uno de sus "diámetros" es 3. Está claro que desde el enfoque de Cooper, si π fuese 3, ¿no sería enteramente natural que las circunferencias fuesen hexágonos? Por mi parte, me intrigó que el "corrimiento" del valor de π no sólo transformase los círculos en hexágonos, sino también el modo interrogativo en modo declarativo. Recordemos que la pregunta pedía decir qué aspecto tendría la propia pregunta en tal extraño mundo subjuntivo. ¿Dejaría acaso la pregunta de sentir curiosidad acerca de sí misma, y por tanto, dejaría de ser una pregunta? No alcanzaba a comprender por qué ese rasgo de la personalidad de la frase habría de quedar afectado por el valor de π. Por otra parte, me parecía obvio que si π *fuese* 3, el antecedente de la condicional ya no estaría en subjuntivo. En efecto, en lugar de decir "si π tuviera el valor 3" debería ahora decir "porque π es 3", o algo por el estilo. Así que reorganizando mis ideas llegué a esta pequeña variación sobre la frase de Cooper: "¿A qué se parece esta frase, siendo π igual a 3 (como es habitual)?"

Varios lectores sintieron interés por frases que aludieran al idioma en que están (o no están, según el caso). Por ejemplo, "Si hablaras en español ahora estarías en tu idioma materno". Jim Propp me envió un delicioso par de frases de este tipo, que es forzoso leer en conjunción:

"Cette phrase se refere a elle-meme, mais de une maniere peu evidente a la plupart des Espagnols."

"Plim glorkle pegram ut replat, trull gen ris clanter froat veb hup lamerack gla smurp terricolas."

En caso de que la primera frase le resulte incomprensible, lo más sencillo puede ser pedirle a uno de sus amigos marcianos que le ayude a descifrar la segunda. Ello le dará algunas claves para traducir la primera. [Pedimos disculpas por la supresión de los acentos ortográficos marcianos, pero carecemos de ellos en esta tipografía.]

El pasado marzo publiqué varias frases creadas por David Moser, mencionando entonces que éste había escrito todo un cuento formado por frases auto-alusivas. Muchos lectores quedaron intrigados. Y por ello he pensado que la mejor manera de cerrar este artículo sería publicar íntegro el cuento de David. Hélo aquí:

Este es el título de este cuento, que se encuentra también varias veces en el cuento mismo

Esta es la primera frase de este cuento. Esta es la segunda frase. Este es el título de este cuento, que se encuentra también varias veces en el cuento mismo. Esta frase pone en duda el valor intrínseco de las dos primeras frases. Esta frase tiene por finalidad informarle, en caso de que no se haya dado cuenta ya, de que éste es un cuento auto-alusivo, es decir, un cuento que contiene frases que aluden a su propia estructura y función. Esta es una frase que da fin al primer párrafo.

Esta es la primera frase de un nuevo párrafo de un cuento auto-alusivo. Esta frase sirve para presentarle a usted el protagonista de este cuento, que es un jovencito llamado Billy. Esta frase está diciéndome que Billy es rubio y norteamericano y que tiene doce años y que está estrangulando a su madre. Esta frase hace un comentario acerca de la peculiar naturaleza de la forma narrativa auto-referida, al mismo tiempo que reconoce el extraño y festivo distanciamiento que concede a su autor. Como para ilustrar la proposición enunciadas en la última frase, esta frase nos recuerda sin traza de ironía que los niños son precioso don divino y que el mundo es lugar mejor cuando se ve agraciado con los gozos únicos que sólo ellos aportan.

Esta frase describe los ojos desorbitados y la lengua protuberante de la madre de Billy y hace referencia a los desagradables ruidos de asfixia y basca que emite. Esta frase hace la observación de que estos son tiempos difíciles e inseguros y que las relaciones personales, incluso las presumiblemente más enraizadas y permanentes, manifiestan positiva tendencia a quebrantarse.

Presenta en este párrafo la técnica del fragmento de frase. Un fragmento de frase. Otro. Buena técnica. Más usado posteriormente.

Esta es en realidad la última frase del cuento y se encuentra aquí por error. Este es el título del cuento, que se encuentra también varias veces en el cuento mismo. Cuando Gregor Samsa se despertó tras una noche de inquieto soñar se encontró a sí mismo yaciendo en su cama y transformado en insecto gigantesco. Esta frase le informa de que la frase precedente ha sido enteramente tomada de otra historia (muy superior, debe observarse) y no tiene lugar en esta narración particular. A pesar de lo declarado en la frase anterior, esta frase se siente obligada a informarle de que la historia que está usted leyendo es en realidad "La Metamorfosis", de Franz Kafka, y que la frase aludida por la frase precedente es la *única* frase que debe verdaderamente encontrarse en este cuento. Esta frase echa por tierra lo anterior, e informa al lector (pobre y confuso infeliz) de que esta pieza literaria es realmente la Declaración de Independencia de los Estados Unidos, pero que el autor, haciendo alarde de la más extrema negligencia (o de malicioso sabotaje) ha sido hasta ahora incapaz de incluir ni siquiera *una sola frase* de tan estimulante documento, si bien ha condescendido a valerse de un pequeño *fragmento* de frase, a saber, "Cuando en el curso de los acontecimientos humanos", inmersa entre comillas hacia el final de una frase. Dando muestra de aguda percepción del aburrimiento y hostilidad del lector medio hacia los fútiles juegos conceptuales de las frases anteriores, *esta* frase nos devuelve por fin al escenario del cuento, formulando la pregunta, "¿por qué Billy está estrangulando a su madre?" Esta frase intenta arrojar algo de luz sobre la pregunta planteada en la frase anterior, pero fracasa. *Esta* frase, no obstante, sí tiene éxito en sugerir la posibilidad de una relación incestuosa entre Billy y su madre y alude a las complicaciones freudianas concomitantes que todo lector astuto puede inmediatamente imaginar. Incesto. El tabú innombrable. La prohibición universal. Incesto. ¿Aprecia los fragmentos de frase? Buen recurso literario. Más usado posteriormente.

Esta es la primera frase de un nuevo párrafo. Esta es la última frase de un nuevo párrafo.

Esta frase puede servir como comienzo o fin de un párrafo, según donde se la coloque. Este es el título de este cuento, que se encuentra también varias veces en el cuento mismo. Esta frase plantea serias objeciones a toda la clase de frases auto-alusivas que se limitan meramente a comentar su propia función o situación en el seno del cuento (por ejemplo, las cuatro últimas frases), fundándose en que son monótonamente predictibles, imperdonablemente complacientes consigo mismas y de que sirven solamente para distraer al

lector del verdadero tema del cuento, que en este punto parece referirse a estrangulación e incesto y quién sabe a qué otros deliciosos temas. El propósito de esta frase es hacer notar que la frase precedente, a pesar de no ser miembro de la clase de frases autoalusivas que tan duramente critica, no sirve empero para otra cosa que distraer al lector del verdadero tema de este cuento, que en realidad trata de la inexplicable transformación de Gregor Samsa en insecto gigantesco (no obstante las vociferantes impugnaciones de otras frases bien intencionadas pero con deficiente información). Esta frase puede servir como comienzo o fin de un párrafo, según donde se la coloque.

Este es el título de este cuento, que se encuentra también varias veces en el cuento mismo. Este es *casi* el título de este cuento, que se encuentra solamente una vez en el cuento mismo. Esta frase enuncia con pesar que hasta ahora el estilo de narración auto-alusiva ha tenido efectos paralizantes sobre el verdadero desarrollo del cuento propiamente dicho, es decir, que estas frases han estado tan ocupadas analizándose a sí mismas y al papel que desempeñan en la narración, que han fracasado rotundamente en llevar a cabo su misión de comunicar los sucesos e ideas que se espera conformen el argumento, desarrollo de personajes, etcétera, en resumen, las *raisons d'être* mismas de cualquier frase respetable y laboriosa con destino definitivo en cualquier pieza de ficción en prosa. Esta señala la evidente analogía entre la apurada situación de estas frases agonizantemente centradas en sí mismas y la de los seres humanos afligidos de modo semejante, y pone de relieve los análogos efectos paralizantes que en éstos provoca la introspección torturada y excesiva.

El objeto de esta frase (que puede también servir de párrafo) es especular que si la Declaración de Independencia hubiera sido estructurada y redactada con igual indolencia e incoherencia que hasta ahora lo ha sido este cuento sería imposible decir en qué clase de sociedad libertina y descarriada estaríamos viviendo en nuestros días o en qué simas de degradación podrían estar sumidos los habitantes de este país que pudieran llegar al extremo de que escritores envilecidos y trastornados construyeran frases exasperante e innecesariamente prolijas e irritantes que poseen a veces la cualidad cuestionable por no decir directamente indeseable de aludir a sí mismas y que a veces carecen incluso de puntuación o manifiestan otras muestras de gramática inexcusablemente torpona como recurrir a redundancias innecesariamente superfluas que casi con certeza podrían tener insidiosos efectos secundarios sobre el estilo de vida y la moral de nuestra impresionable juventud, induciéndola a cometer incesto e incluso el asesinato y tal vez sea ésta la causa de que Billy esté estrangulando a su madre, a causa de *frases como ésta*, que carecen de objetivo discernible o finalidad inteligible y que acaban en cualquier sitio, incluso en su mit.

Chocante. Un fragmento de frase. Otro fragmento. Edad, doce años. Esta es una frase que. Fragmentada. Y estrangulando a su madre. Lo siento, lo siento. Chocante. Esta. Más fragmentos. El título de este cuento, que. Rubio. Lo siento, lo siento. Fragmento tras fragmento. Más duro. Esta es una frase que. Fragmentos. Recurso condenadamente bueno.

Esta frase tiene un triple propósito: (1) pedir disculpas por el desdichado e inexplicable lapsus que salta a la vista en el párrafo anterior; (2) asegurarle a usted, el lector, que eso no volverá a suceder y (3) reiterar lo dicho en el sentido de que estos son tiempos inciertos y difíciles y que ciertos aspectos del lenguaje, incluso los presumiblemente más enraizados y permanentes como la sintaxis y la significación, llegan a quebrantarse. Esta frase no añade nada sustancial a los sentimientos de la frase anterior, limitándose a proporcionar frase final a este párrafo, que de lo contrario podría carecer de ella.

Esta frase, en un repentino y valeroso alarde de altruismo, se propone abandonar el modo auto-alusivo, pero fracasa. Esta frase lo intenta también, pero su esfuerzo está condenado desde su nacimiento.

Esta frase, luchando hasta el último cartucho para infundir una pizca de argumento en esta paralítica pieza en prosa, hace rápida alusión a los frenéticos esfuerzos de Billy para encubrir su acción, seguida de un pasaje lírico, conmovedor y delicadamente escrito en el cual Billy se reconcilia con su padre (resolviendo así los conflictos freudianos subliminales evidentes para cualquier lector astuto), y una excitante persecución policíaca final durante la cual Billy es accidentalmente muerto de un tiro por un policía bisoño y dominado por el pánico, que por coincidencia también se llama Billy. Esta frase, aunque fundamentalmente henchida de simpatía hacia los laudables esfuerzos de la última frase tan rica en acontecimientos, sirve para recordar al lector que tales alusiones a una historia que, de hecho, todavía no existe pueden aceptarse como sustitutos de la cosa genuina, y por consiguiente no bastarán para descolgar al autor (holgazán chapucero e indolente) del gancho proverbial.

Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo. Párrafo.

El propósito. De este párrafo. Es pedir disculpas. Por su gratuita utilización. De fragmentos. Lo siento.

El propósito de esta frase es pedir disculpas por los fútiles y estultos juegos de adolescente en que incurren los dos párrafos precedentes y expresar la consternación que por nuestra parte, nosotras, las frases de mayor madurez, sentimos a causa de que el tono de este cuento sea tal que parezca incapaz de comunicar una sencilla, aunque no poco sórdida, trama narrativa.

Esta frase desea pedir disculpas por todas las innecesarias disculpas solicitadas en este cuento (incluidas éstas), que aunque situada aquí en beneficio de los lectores más molestos, en realidad sólo pospone por métodos enloquecedoramente iterativos la continuación de la línea narrativa a punto de caer en el olvido.

Esta frase revienta por sus signos ortográficos con la noticia de las horribles implicaciones de la auto-alusión aplicada a frases, práctica que muy bien pudiera resultar una auténtica caja de Pandora generadora de potencial devastación, pues si una frase puede referirse o aludir a sí misma ¿por qué no habría de poder hacerlo igualmente una plebeya cláusula subordinada, *esta misma* cláusula tal vez? ¿O este fragmento de frase? ¿O tres palabras? ¿Dos palabras? ¿*Una*?

Tal vez convenga que esta frase, suavemente y sin traza de condescendencia, nos recuerde que éstos son tiempos verdaderamente muy difíciles e inseguros, y que, en general, las personas sencillamente no son lo bastante correctas unas con otras, que tal vez nosotros, tanto como seres humanos conscientes como cuanto frases conscientes, tengamos que *esforzarnos más*. Quiero con esto decir que *hay* algo llamado libre albedrío, *tiene* que haberlo, ¡y esta frase es prueba de ello! Ni esta frase ni usted, el lector, se encuentran completamente desasistidos frente a las fuerzas inmisericordes que operan en el universo. Debemos permanecer firmes en nuestro suelo, encarar los hechos, asir fuertemente a la Madre Naturaleza por el cuello y *esforzarnos más*. Por el cuello. Más. Esforzarnos. Más.

Lo siento.

Esta es la última frase del cuento. Esta es la última frase del cuento. Esta es la última frase del cuento. Esta es.

Lo siento.

Taller y laboratorio

¿Por qué se mantienen unidas las partículas de arena y barro cuando se humedecen?

Jearl Walker

Esos castillos de arena y flanes de barro que suelen construirse como pasatiempo plantean algunas cuestiones intrigantes. Aunque la arena y el barro están formados por partículas cuya composición es casi la misma, presentan propiedades muy diferentes. ¿Cuáles son las fuerzas que impiden desmoronarse tales obras de ingeniería recreativa? ¿Por qué caen los castillos al secarse, pero no los flanes de barro? ¿A qué obedece que los flanes se contraigan al secarse y los castillos de arena no? ¿De qué modo el agua, que suele actuar de lubricante, aumenta la cohesividad de la arena y arcilla? Las respuestas a estas preguntas se encuentran en las interacciones eléctricas entre el agua y las partículas de arena o arcilla.

La arena se compone de granos más o menos esféricos y relativamente grandes, que se mantienen siempre muy apretados unos contra otros; incluso estando húmedos, como revela el hecho de que no se contraiga cuando pierde agua. En la superficie de los granos humedecidos hay iones positivos y negativos, habitualmente emparejados y con uno de los miembros del par más separado de la superficie que el otro. El ion que aparece en la posición más externa no es el mismo en todos los puntos de la superficie. Ambas disposiciones contribuyen a crear un campo eléctrico en el espacio que rodea a cada grano. Cuando el ion externo es positivo, el vector que representa el campo eléctrico apunta hacia el exterior de la superficie; cuando el ion externo es negativo, dicho vector señala hacia el interior de la superficie.

Aunque pueda pensarse que el campo eléctrico medio que rodea a cada grano sea nulo, no es así; la razón yace en que la carga de los iones negativos contiguos a la superficie está polarizada. A uno de los lados de cada ion negativo se sitúa un ion positivo, algo más separado de la superficie, mientras que al otro lado del ion negativo no hay carga alguna. Por ser la carga negativa del ion negativo superior a su carga positiva, la nube de electrones se desplaza ligeramente, separándose del núcleo, cuya carga es positiva, en dirección al ion positivo vecino. Esta ligera separación de cargas positivas y negativas constituye lo que se llama dipolo eléctrico y produce un campo eléctrico débil dirigido hacia el exterior del grano de arena. Por su parte, los iones positivos en contacto con la superficie son deficitarios en carga negativa, sus cargas no se separan ni un ápice y crean el campo eléctrico de una carga puntual y no el de un dipolo, que es más débil. Por todo ello, el campo eléctrico alrededor de un grano de arena se debe principalmente a los iones positivos en contacto con la superficie del grano.

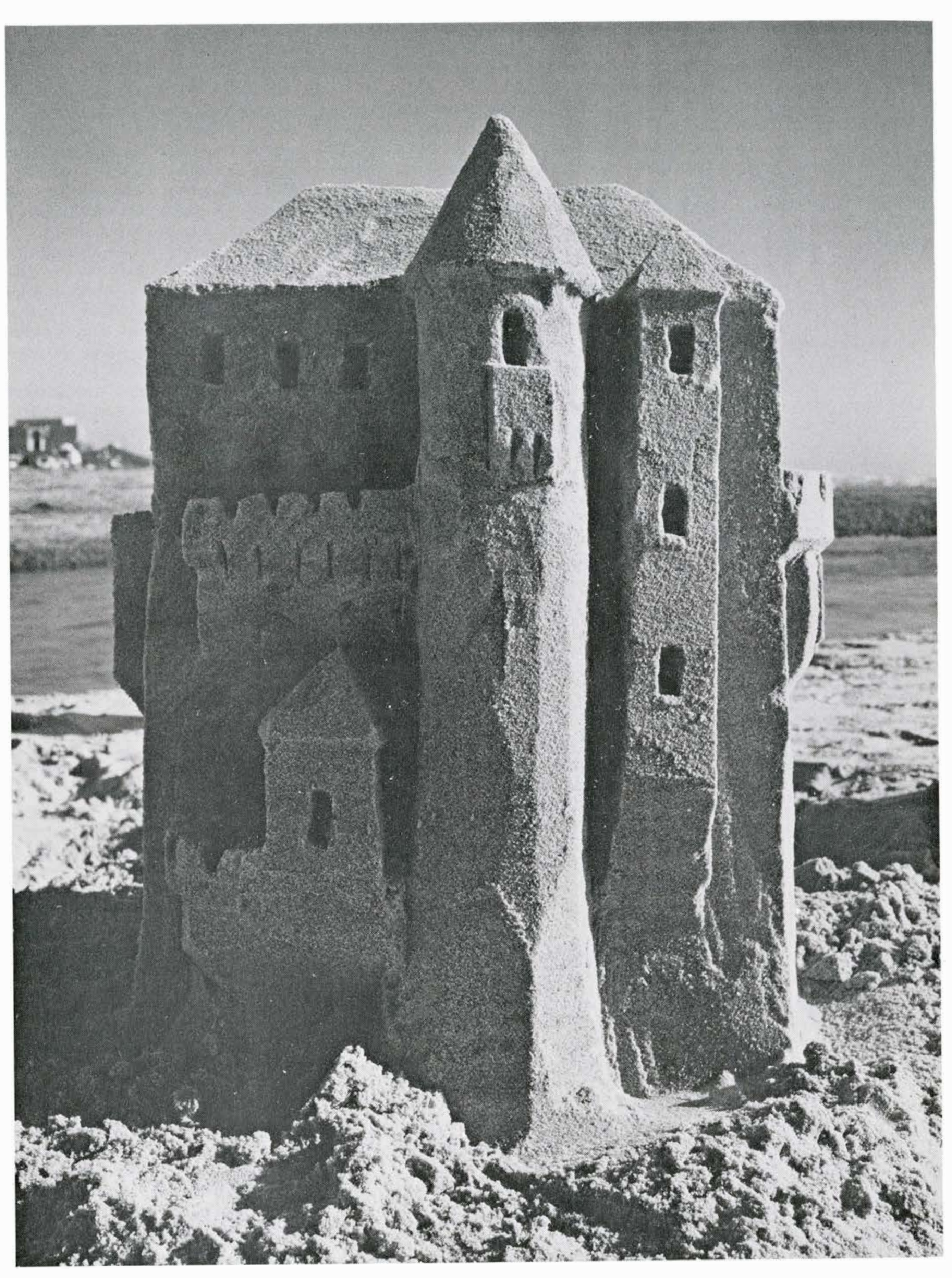
La arena húmeda es estable y consistente a causa de la interacción eléctrica entre dichos iones positivos y el agua intersticial. Según un modelo un poco anticuado de esta interacción, el agua proporciona iones negativos (OH^-) que se adhieren a los granos y apantallan eléctricamente los lugares de carga positiva. Por otra parte, los iones positivos (H_3O^+) permanecen dentro del agua en puntos más alejados de la superficie. Así, la atracción entre los iones positivos en el seno del agua y los negativos adheridos a la superficie de los granos dificulta el resbalamiento de unos granos sobre otros. Este modelo, aunque desde luego explica la estabilidad de la arena húmeda, presenta un fallo: no es verosímil que el agua se ionice hasta el punto de proporcionar iones OH^- y H_3O^+ en cantidad suficiente.

Ninguno de los modelos actuales para la arena húmeda es completo, pero hay una interpretación moderna de la interacción eléctrica que no requiere que el agua esté ionizada. Aquí se admite que el campo eléctrico generado por los lugares de la superficie del grano positivamente cargados reduce, en parte, la carga positiva del agua contigua a la superficie. Esta reducción no es una ionización, pues las cargas interiores de cada molécula de agua no se separan del todo. Simplemente el campo eléctrico desplaza las posiciones medias de los núcleos de hidrógeno (protones) en las moléculas más alejadas de la superficie. La capa vecina a la superficie queda relativamente negativa, mientras que el agua más alejada de la misma es relativamente positiva, de modo que el paso de agua relativamente negativa a agua relativamente positiva es gradual. Esta situación se define como capa de carga doble difusa.

En el modelo iónico primitivo, la solidez de la arena húmeda se atribuía a la atracción entre los iones de carga contraria existentes en el agua. En el modelo de la capa doble difusa, la solidez tiene su origen en el hecho de que el agua tiende a inmovilizarse. En este caso, como los campos creados por los lugares cargados positivamente de cada grano de arena redistribuyen las posiciones medias de los protones de hidrógeno de agua, se reduce la aptitud de ésta para fluir. La arena es entonces consistente porque el aumento de viscosidad del agua entorpece el deslizamiento de unos granos respecto a otros.

La inmovilización del agua disminuye con la distancia a la superficie de los granos. Por ello, cuando la arena húmeda se hace fluida hasta el punto de no servir para construir castillos de arena, ello se debe a la gran cantidad de agua que hay entre los granos. Y, aunque el agua relativamente negativa alrededor de cada grano siga atrayendo al agua relativamente positiva más alejada, los granos están demasiado distanciados por causa del agua, parte de la cual no estará lo bastante cercana a un grano para participar de un modo significativo en la interacción eléctrica. Será ésta un agua menos viscosa que permitirá el mutuo resbalamiento de los granos (con sus aureolas de agua relativamente cargada).

Es idea muy extendida que los casti-



De **Sandcastles** *(1981), de Joseph Allen, Don McQuiston, Debra McQuiston y Marshall Harrington, editado por Doubleday, Inc.*

llos de arena no se desmoronan gracias a la fuerza ejercida por la tensión superficial del agua intersticial. Se dice que las moléculas de agua son polares porque cada una posee un dipolo eléctrico permanente. Estos dipolos se atraen entre sí. En una superficie de separación entre aire y agua la atracción mutua entre las moléculas de agua vecinas pone a dicha superficie virtualmente en estado de tensión. Cuando la superficie es curva, dicha atracción mutua crea una fuerza resultante que actúa sobre ella.

Sobre la superficie de un castillo de arena tal fuerza resultante favorece el afianzamiento de la arena. Pero, en el interior del castillo, donde no hay huecos con aire, la arena se encuentra igualmente afianzada. Aunque las moléculas se atraen entre sí débilmente, merced al campo eléctrico de sus dipolos eléctricos, tal atracción, poco intensa, no puede ser responsable del afianzamiento y resistencia a fluir de la arena. Igualmente, en un vaso de agua, aunque las moléculas de ésta se atraigan entre sí, el agua no está afianzada.

Cuando nuestra “obra” se seca, se deshace porque su contenido de agua relativamente cargada ha disminuido. En el intermedio de transición, las cargas que hay en la superficie de los granos tornan a distribuirse, formando los iones una capa de carga doble, llamada capa doble de Helmholtz, en la que los iones positivos se sitúan ligeramente más al interior de la superficie y debajo de los iones negativos, que se sitúan exteriormente. El campo eléctrico creado por esta distribución de las cargas es débil, salvo en la misma superficie. Cuando dos granos secos se tocan entonces, se repelen: ambas superficies poseen una capa externa de iones negativos; tal es la razón de que los granos de arena secos no se adhieran unos a otros.

La cohesión de la arena se destruye si sustituimos parte del agua por glicerina. Las moléculas de ésta apantallan eléctricamente los lugares de carga positiva de la superficie de los granos, reduciendo el campo eléctrico en el seno del agua. Al ser más débil el campo eléctrico, las capas de agua se organizan menos y, al no inmovilizarse el agua, la arena no se mantiene ya unida.

El propio lector puede comprobar el comportamiento de la glicerina. Rellene con arena húmeda un pequeño recipiente, apelmazándola hasta que se encuentre bastante compacta. Luego coloque una placa plana sobre el recipiente y, rápidamente, dé la vuelta a placa y recipiente sin permitir que la arena se desparrame. Con la placa sobre el suelo, tire del recipiente poco a poco hacia arriba, de modo que en la placa quede un montículo de arena. Nos convenceremos de la solidez del montículo sacudiendo o golpeando ligeramente la placa.

Se repite el ensayo vertiendo glicerina al agua que se mezcla con la arena. Una vez dada la vuelta al conjunto, tírese del recipiente, hacia arriba de nuevo. Lo más probable es que el montículo que queda sea notablemente más débil y que una sacudida o un leve golpe hagan que se venga abajo. Cuanto más glicerina se añada a la arena, tanto menos estable será el montículo. Este hecho ofrece una paradoja inesperada: cuando se introduce un fluido viscoso (glicerina) en una mezcla muy viscosa (arena con agua), la viscosidad disminuye.

Modelo iónico del agua

La diferencia entre la arcilla y la arena húmedas presenta varios aspectos sutiles. Las partículas de arcilla son considerablemente más pequeñas que los granos de arena y tienen forma de placa, siendo una de sus dimensiones mucho menor que las otras dos; si la arcilla está humedecida, las partículas andan muy sueltas y muy separadas por el agua. Un indicio de su friabilidad es que la arcilla se contrae al secarse.

En la superficie de las partículas de arcilla quedan al descubierto los lugares cargados positivamente, al igual que en los granos de arena. Según el primitivo modelo iónico para la arcilla húmeda, a dichos lugares se adhieren iones OH^-, abandonando iones H_3O^+ en el agua intersticial; la solidez de la arcilla húmeda débese, pues, a la atracción entre ambas clases de iones. En el modelo más moderno para la arcilla húmeda se afirma que el agua se inmoviliza en virtud de la acción del campo eléctrico creado por los lugares cargados positivamente, existentes en la superficie de las partículas. Como en el modelo para la arena húmeda, el agua inmovilizada posee una viscosidad elevada y la arcilla húmeda es estable y consistente porque sus partículas no pueden fluir unas con respecto a otras. Pero si la arcilla está muy húmeda, ya no será así, pues el agua más alejada de las partículas poseerá menos viscosidad y podrá fluir.

Arcilla y arena presentan una notable diferencia en su proceso de secado. Mientras que las capas en torno a los granos de arena poseen movilidad y pueden originar una capa doble de Helmholtz, en las partículas de arcilla no ocurre igual. Ello obedece a que estas últimas son demasiado aplastadas para que puedan formarse capas dobles en las caras opuestas, de mayor tamaño, pues las capas de carga más profundas (que tendrían el mismo signo) se repelerían muy intensamente y no serían estables.

Al secarse la arcilla húmeda, el hecho de que no pueda formarse una capa doble de Helmholtz obliga a las partículas a retener su capa doble difusa y a que permanezca entre ellas el agua viscosa resultante. Así, la estructura de la arcilla sigue siendo estable y consistente; la verdad es que aumenta su rigidez, ya que las partículas se aproximan mutuamente incrementando aún más la inmovilización del agua. A medida que la estructura se contrae, y aparecen más puntos de contacto entre las partículas, éstas se traban compartiendo iones. Muestran, además, resistencia al movi-

miento a causa de la viscosidad del agua, al tiempo que otras fuerzas, de naturaleza eléctrica todas, coadyuvan asimismo a la adherencia.

Los granos de arena se aprietan cerradamente unos contra otros. Pero las partículas de arcilla andan mucho más sueltas. Diferencia que es secuela de sus distintos tamaño y forma. Consideremos un grano que cae y se pone en contacto con otro en reposo. ¿Quedará adherido a éste, rodará o resbalará sobre el segundo para proseguir su movimiento? En la zona de contacto, las fuerzas son suficientemente intensas para mantenerlos unidos, pero la atracción gravitatoria sobre el grano móvil somete a éste a un momento de giro que tiende a hacerlo girar en torno al punto de contacto con el otro grano. Aunque algunas de las fuerzas atractivas crean entonces un momento de giro opuesto a esa rotación, como el grano es relativamente pesado, el momento de giro debido a la gravedad vence al debido a las fuerzas atractivas, y el grano rota. Este movimiento continúa haciendo que el grano descienda sobre los demás hasta que quede bien apretado contra ellos.

¿Qué ocurre cuando una partícula de arcilla cae sobre un montón de partículas de arcilla? Se ve sometida a fuerzas atractivas de la partícula inmóvil que le sale al paso. La gravedad continúa atrayéndola hacia abajo, creando un momento de giro que tiende a hacerla rotar en torno al punto de contacto. Pero la partícula es tan liviana que el momento de giro debido a la gravedad no puede llevar a cabo la rotación y, por ello, la partícula permanece con la misma orientación inicial que tenía al chocar con la partícula inmóvil. Cuando caen más partículas, éstas tampoco avanzan más, una vez que aterrizan. De lo que resulta que las partículas se agrupen con holgura.

Comprimamos arena seca o húmeda. Esta cambiará de forma sólo en el caso de que los granos de arena resbalen unos respecto a otros. Aplíquese una leve presión a la arena húmeda, insuficiente para desencadenar el movimiento de unos granos respecto a otros. Entonces, la superficie de contacto de cada pareja de granos que se aprieten uno contra otro aumentará, al igual que la intensidad de las fuerzas eléctricas que los mantengan en contacto. En el centro de la superficie de contacto, la presión adicional ejercida sobre los granos deformará, irreversiblemente, la superficie de los mismos. La energía gastada en esa deformación no se recuperará al retirar la presión sobre los granos. Sin embargo, la mayor parte de la energía de compresión de los granos servirá para deformar la zona que rodee el centro de la superficie de contacto, proceso éste no irreversible. Esta última energía se almacenará en forma de energía potencial elástica, lo mismo que la energía que se almacena en un resorte comprimido.

Cuando se retira la presión sobre los granos, la elasticidad de éstos los impulsa a separarse. Además, las fuerzas eléctricas que se generan durante el tiempo que dura el mayor contacto entre los granos son relativamente débiles y el rebote las vence sin dificultad. Por consiguiente, cuando se retira la presión los granos retornan a su forma primitiva.

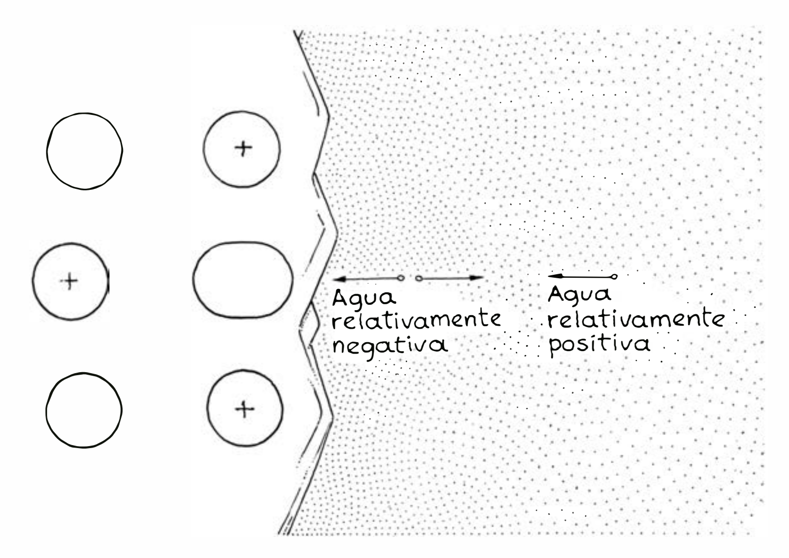
Comprimamos ahora arcilla húmeda. Las cosas cambiarán algo. Consideremos el montón de partículas de arcilla sueltas representado en la ilustración de la página 119. En ella, la fuerza aplicada flexiona a una de las partículas, que son aplanadas, con lo que el extremo antes libre de la misma entra en contacto con otra partícula. En el punto de contacto, las fuerzas atractivas hacen que ambas partículas se traben. Cuando se retira la fuerza aplicada exteriormente, las partículas siguen trabadas, porque las fuerzas atractivas son suficientemente intensas para mantenerlas así. Por ello, al apretar sobre arcilla húmeda con la punta de un dedo, lo que se hace es comprimir esa estructura, que no rebotará si cesa la compresión. De estos materiales se dice que son plásticos. La arcilla húmeda es plástica; la arena húmeda, no.

Las mejores arenas para construir castillos son las de bajo contenido de arcilla. Es preferible la arena próxima al agua, que ya se encuentre bastante húmeda; pero no la que baña el oleaje, porque ésta, humedecida en exceso, no resultaría consistente. Además, la marea destruiría enseguida un castillo levantado en la misma linde.

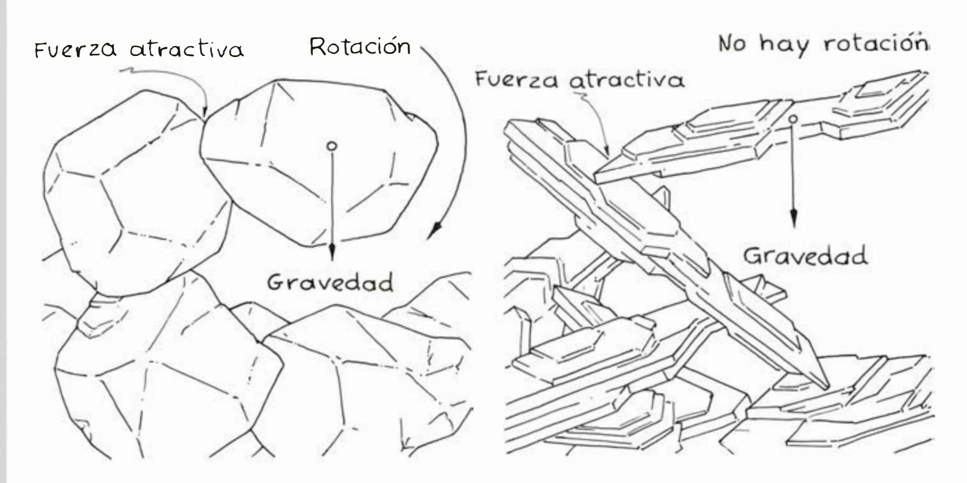
Hay varios procedimientos para construir castillos de arena. Así, puede volcarse cuidadosamente un recipiente lleno de arena húmeda, de modo que se conserve compacta y guarde la forma interior del molde. Otro método consiste en conformar la arena apretándola entre dos trozos de cartón, lo que puede proporcionar unas murallas de buena longitud. Técnica más difícil es la de escultura, que consiste en formar un montículo y luego, empezando por arriba, retirar material poco a poco. Hay un método que se inspira en el crecimiento de las estalagmitas de las cavernas; desde el puño cerrado lleno de arena húmeda se deja caer ésta lentamente en forma de chorro delgado, y así se crea una delgada aguja semejante al campanario de una catedral futurista.

Si queremos que nuestra obra dure todo un día de verano, verteremos agua de tanto en tanto, al objeto de preservar su consistencia. El agua añadida se infiltra entre los granos, donde los campos eléctricos que éstos han creado desplazan las posiciones medias de los protones de hidrógeno del agua y así la inmovilizan. El aumento de viscosidad del agua asegurará su estabilidad hasta que se haga necesaria otra rociada.

Aunque la arena no cambie de volumen el secarse, sí se dilata cuando sufre un esfuerzo de compresión brusco, fenómeno que recibe el nombre de dilatancia. Quizás el lector lo haya observado paseando por un trecho de playa



Inmovilización del agua



*Apilamiento de la arena (***izquierda***) y de la arcilla (***derecha***)*

recién bañado por una ola. Entonces la arena parece húmeda antes de pisarla, pero se seca en cuanto ponemos el pie y, para que la huella vuelva a parecer húmeda, ha de transcurrir un rato.

La arcilla húmeda no se seca bajo las pisadas porque es friable; cualquier presión que sufra sirve para que las partículas se acerquen mutuamente y disminuya de volumen. La dilatancia es característica de los materiales cuyos granos se aprietan mucho unos contra otros; de tal modo que, cuando reciban una compresión repentina que les obligue a desplazarse, sólo podrán hacerlo a costa de aumentar su separación media. Por tanto, cuando la arena húmeda reciba el pisotón, la dilatación consiguiente implica que la superficie de la arena ha de elevarse, dejando el agua por debajo; y ésta necesitará dos o tres minutos para infiltrarse hasta la superficie.

En un alud de arena seca participa también la dilatancia. Para conocer las propiedades más importantes de un alud basta diseñar un experimento muy sencillo. Inclinemos gradualmente una artesa con arena en su interior, hasta que comience a fluir. El alud se inicia cuando la inclinación de la superficie supera el llamado ángulo de rozamiento dinámico. Con el resbalamiento de la arena dicha inclinación va disminuyendo hasta el llamado ángulo de rozamiento, para el que la superficie es estable. Si la artesa vuelve a inclinarse, la superficie de la arena volverá a superar el ángulo de rozamiento dinámico y se iniciará otro alud.

Para todo ángulo inferior al de rozamiento, los granos de arena que se encuentran en la superficie inclinada permanecen inmóviles, al estar muy apretados unos contra otros, y sólo deslizarán si la inclinación de la pendiente alcanza el ángulo de rozamiento dinámico. Para que esto ocurra, los granos primero deben elevarse de modo que puedan resbalar o rodar sobre los granos situados por debajo. Entonces, toda la sobrecarga de arena entra en dilatancia y se desliza pendiente abajo.

Gran número de sustancias granulares se deslizan prácticamente de la misma manera. La nieve es un caso bien conocido. Además de la arena, me he ocupado del azúcar (granular y en polvo), la sal, el almidón y el cacao en polvo; también he estado estudiando los aludes de alubias pintas secas, naranjas y manzanas. En cada caso existen unos ángulos de rozamiento y de rozamiento dinámico particulares.

En el caso de la arena, la diferencia entre ambos ángulos se encuentra en torno a los cinco grados. Si la arena se rocía con agua, los granos ya no resbalan hacia abajo por la superficie inclinada; en lugar de ello, toda la masa de arena desliza por el fondo de la artesa cuando la misma se inclina convenientemente. Para las alubias pintas, la diferencia entre los dos ángulos es de unos dos grados. El almidón y el cacao en polvo son tan cohesivos que en su caso encontré dificultades para realizar las mediciones.

Hay muchos polvos que son cohesivos; o sea, están constituidos por partículas que se pegan unas a otras y se resisten a fluir. Algunas, como los polvos faciales, se fabrican precisamente para ser cohesivas y mantenerse adheridas. En otras, como el almidón y el cacao en polvo, la cohesión estorba. La cohesión en los polvos se debe a fuerzas eléctricas de naturaleza diversa.

Una de tales fuerzas, la electrostática, se debe a las cargas superficiales existentes en partículas contiguas. La fuerza de Coulomb es la más intensa de

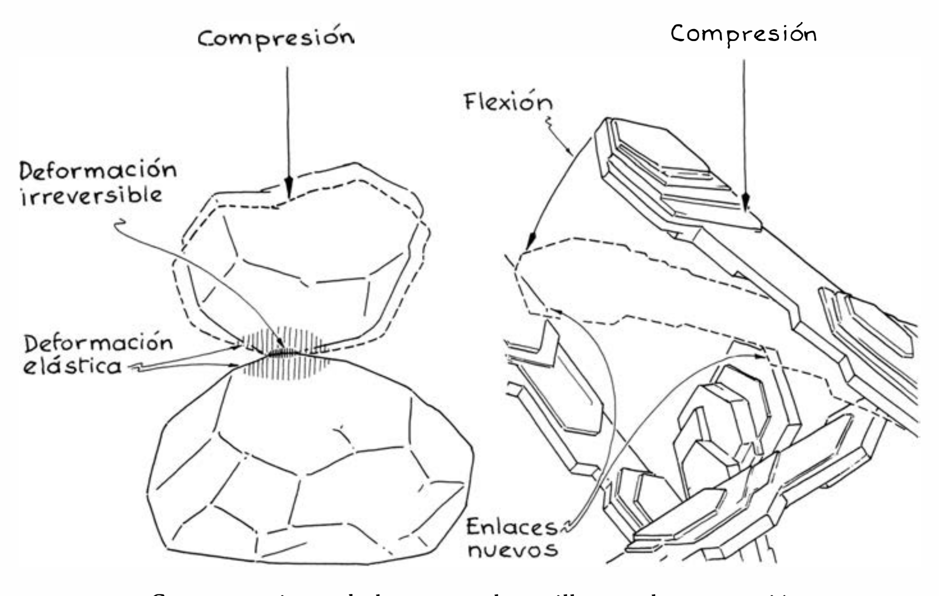
las fuerzas electrostáticas, en virtud de la cual se repelen las partículas cargadas con el mismo signo y se atraen las cargadas con signos contrarios. Más débil es la fuerza que se genera entre las partículas cargadas y los átomos y moléculas neutros (o sea, sin carga). En este caso, el campo eléctrico de las partículas cargadas provoca la separación de las cargas de los átomos y moléculas; por ejemplo, si una partícula es negativa, la componente positiva del átomo o molécula se desplaza ligeramente hacia aquélla, de manera que, aunque el átomo o molécula siga siendo positivo, éste será entonces un dipolo eléctrico atraído hacia la partícula cargada.

Ciertas moléculas son dipolos eléctricos permanentes que, por ello, se atraen unas a otras cuando están próximas; el extremo negativo de una tira hacia el extremo positivo de otra. La fuerza de van der Waals es una fuerza eléctrica de tipo parecido, aunque se debe a la naturaleza cuántica de los átomos. Si bien un átomo puede ser neutro eléctricamente, y así no constituir un dipolo eléctrico, puede serlo incluso en cada instante dado. Efectivamente, en el modelo del átomo de Bohr se supone que el electrón orbita alrededor del núcleo; por tanto, siempre se encuentra en el borde del átomo y, consiguientemente, separado de la carga positiva, constituyendo un dipolo instantáneo capaz de inducir un dipolo en cualquier átomo vecino. Una vez que se establecen ambos campos dipolares, los átomos se atraen.

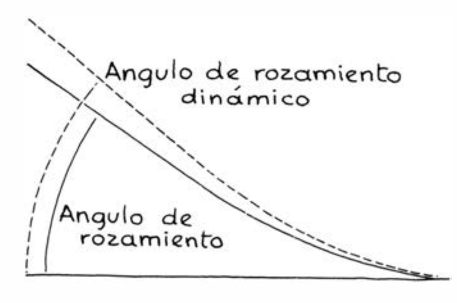
La fuerza de van der Waals insta la atracción entre dos superficies en contacto. Y si los objetos materiales no se aproximan es sólo porque interviene la repulsión de Coulomb que lo impide. Muchos de los polvos más comunes son cohesivos merced a la fuerza de van der Waals.

Cuando las partículas de dichos polvos se tocan, las fuerzas de rozamiento que se desarrollan entre ellas impiden su deslizamiento relativo. Si entre las partículas hay agua, su cohesión puede obedecer a la inmovilización de la misma por acción de los lugares cargados de las partículas. La tensión superficial puede asimismo desempeñar un papel si hay aire entre las partículas, o si el exterior del terrón del polvo en cuestión se mantiene unido mediante agua.

La cohesión de un producto pulverulento no puede predecirse a partir de un conocimiento detallado de la estructura de sus partículas, salvo quizá de una manera muy general. Quizás el rozamiento sea mayor entre partículas irregularmente conformadas, pero sólo si llegan a trabar contacto. Así es, ya que la cohesividad de una sustancia pulverulenta suele reducirse cuando las partículas se muelen y adquieren una conformación más lisa y regular. También puede alterarse la cohesividad desecando el producto al objeto de rebajar la influencia del agua intersticial. Lo corriente es que aumente la cohesividad de un material pulverulento al comprimirlo, porque las fuerzas de van der Waals y electrostática se hagan más intensas o porque crezca el número de partículas en contacto o trabadas. En general, cuanto más pequeños son los granos de un polvo, tanto más cohesivo es éste. Supóngase que preparamos un producto pulverulento, el tamaño de cuyas partículas sea casi uniforme, y que lo introducimos en una tolva dota-



Comportamiento de la arena y la arcilla ante la compresión



Angulos que intervienen en los aludes

da de una apertura circular en el fondo que le permita descargarse. Para medir la velocidad de descarga se mide el tiempo necesario para que de la tolva salga una determinada cantidad de polvo.

Cuando las partículas son muy pequeñas, el régimen de descarga es muy lento, o nulo, pues las fuerzas cohesivas son intensas y limitan su movimiento relativo. Si estas partículas se van sustituyendo paulatinamente por otras de mayor tamaño, aumenta el régimen de descarga hasta alcanzar un valor máximo, para luego decrecer; de modo que cuando el tamaño de las partículas es, más o menos, una quinta parte del tamaño de la abertura de salida, la descarga cesa. En ese momento, las partículas se habrán anclado unas en otras formando, a través de la abertura, algo similar a un muro de piedra en seco.

Hay productos pulverulentos que se apelmazan en grumos, actuando el agua intersticial de consolidante. Por su forma, las capas de agua intersticial se dividen en tres categorías. Si, inicialmente, las partículas están muy secas, el agua que hay entre los puntos de contacto de las partículas se encuentra en lo que se llama formación colgada, donde existen zonas intersticiales llenas de aire. Si las partículas no lo están tanto, son más extensas las zonas llenas de agua, pero siguen separadas por zonas de aire; esta formación se llama funicular. Cuando los intersticios se llenan por completo de agua, la formación se llama capilar. No se conoce bien la naturaleza pormenorizada de las fuerzas que mantienen unidos los grumos. La adherencia suele atribuirse a la tensión superficial del agua; esta hipótesis puede que baste para las formaciones colgada y funicular. En todo caso, la inmovilización del agua es importante en cualquiera de las tres formaciones.

Los polvos pueden apelmazarse, aunque estén casi secos. Quizás el lector ya haya advertido que, al agregar agua al chocolate en polvo, muchas veces éste se encuentra tan apelmazado que se resiste a disolverse y forma unos terrones secos por dentro y húmedos por fuera. Cuando en la superficie del agua flota una bola gruesa de chocolate en polvo, la parte que está en el borde del agua hace que se desmorone el perímetro del grumo, lanzando al agua un alud con parte del polvo.

Los polvos se apelmazan cuando el agua intersticial de la parte externa de un terrón es tan viscosa (a causa de la inmovilización) que al agua le cuesta infiltrarse en el interior. Aquí puede que también intervengan las fuerzas atractivas entre partículas. Pero, finalmente, en los intersticios acaba infiltrándose agua suficiente, con lo cual el campo eléctrico creado por las partículas no puede ya inmovilizar toda el agua, ni aquellas atraerse con fuerza. El terrón se desmorona. Aun así, quedan partículas que permanecen agrupadas largo rato tras la disolución y desaparición de la masa principal de polvo.

La fécula húmeda se comporta de un modo similar a otros muchos productos pulverulentos humedecidos; su viscosidad depende, empero, de la compresión a que esté sometida. Se trata de un caso de fluido no newtoniano, de cuyas propiedades traté aquí en enero de 1979. La fécula es tixótropa; es decir, su viscosidad aumenta con la compresión.

Desleí ligeramente con agua un poco de fécula, la arrollé en forma de pelota y luego la examiné periódicamente mientras se secaba. Al principio, la solidez de la bola dependía de la compresión que le ejercía: al presionarla se presentaba muy sólida y parecía seca. Si aflojaba la compresión, la bola fluía inmediatamente y parecía húmeda. A la vez que el agua se iba evaporando, la bola adquiría mayor consistencia, hasta que su superficie se tornó rígida al tacto. Aquella estructura no resistía esfuerzos de torsión, pero se mostraba rígida a la presión.

Como la fécula se compone de partículas muy pequeñas, para su comportamiento podría establecerse un modelo basado en las interacciones eléctricas dentro de la arcilla húmeda. No sé muy bien por qué la fécula es tixótropa. La explicación más obvia sostiene que el aumento de viscosidad está ligado a la inmovilización del agua intersticial. Así, al comprimir una bola de fécula, las partículas de ésta se acercan mutuamente y disminuye la movilidad del agua intersticial, al quedar ceñida por las partículas; de esa manera, la zona en cuestión de la bola se torna mucho más viscosa que antes de aplicar la presión.

La mayoría de los productos pulverulentos cohesivos son, además, adhesivos: se pegan a los dedos y a los utensilios de cocina. En lo que respecta a la adherencia, tal vez resulte difícil determinar cuáles son las fuerzas eléctricas importantes, de entre las muchas posibles. Las de van der Waals intervienen habitualmente, aunque cabe presumir la participación de algunas de las fuerzas electrostáticas. Si las partículas están húmedas y son más bien grandes, pudiera ser que la tensión superficial explicase suficientemente la adherencia.

Los polvos finos de grano uniforme captan un poco de humedad ambiental y parte de su atracción por los dedos se debe, supongo, a la inmovilización del agua entre los dedos y las partículas del polvo. Efectivamente, en la superficie de cada partícula hay una capa de agua relativamente negativa y, algo más separada, hay agua relativamente positiva. Entonces, cuando ese sistema se acerque a una superficie, la de un dedo por ejemplo, el campo eléctrico del agua relativamente positiva obligará a la superficie a hacerse negativa. Así, cada partícula se pega al dedo en virtud de la atracción de Coulomb entre capas de cargas contiguas: la superficie de la partícula es positiva, el agua contigua a ésta es relativamente negativa, el agua más separada de la partícula es relativamente positiva y la superficie del dedo es negativa.

Este tipo de adherencia, empero, sólo es intensa cuando el tamaño de las partículas ronda el orden de pocos micrometros; es casi seguro que las partículas de mayor tamaño resulten demasiado pesadas para quedar adheridas a los dedos. Tampoco se adhieren las partículas mucho más pequeñas, mas no por razón de su peso. En este caso, lo que ocurre es que las partículas más pequeñas crean unas aureolas de agua inmovilizada de mayor espesor, al no ser tan aptas para generar capas dobles de Helmholtz. De este modo no pueden acercarse tanto a los dedos como las partículas algo mayores, cuyas aureolas acuosas son más finas, y se desprenden. Afortunadamente, la adherencia a los dedos es así de selectiva; de lo contrario, nuestra vida se haría bastante incómoda, especialmente en los días húmedos, cuando cualquier cosa que tocáramos se nos pegase a los dedos por mor de la inmovilización del agua existente entre ellos y las partículas.

Libros

Gen y cultura, filosofía social, Ray lexicógrafo y terremotos

Ramón Margalef, Emilio Lledó, Luis Alonso y Antoni M. Correig

GENES, MIND, AND CULTURE: THE COEVOLUTIONARY PROCESS, por Ch. J. Lumsden y E. O. Wilson. Harvard University Press, Cambridge, Mass. y London, 428 págs. 1981. Edward Wilson, distinguido estudioso de los insectos sociales, publicó, en 1975, un volumen con el título *Sociobiology: The new synthesis*, del que, en su día, publiqué un comentario en la "Revista de Occidente" que, por los azares de las imprentas o de los compaginadores, salió sin referencia expresa al libro. Por cierto que esta obra es ya asequible en español, editada por Omega, de Barcelona. El tema central de dicho libro era que la cultura tiene una base genética, en lo que será difícil encontrar un biólogo que no esté de acuerdo; pero el libro mencionado estaba escrito de manera que, además de hacer pensar, podía enfurecer a muchos lectores. En mi comentario escribí que, según fueran los lectores, verían al libro como "materialista ateo" o como "elitista y fascista" y, en cualquier caso, lo considerarían subversivo. Así ocurrió, y el libro, sus escritos posteriores y el propio Wilson han sido centro de polémicas vivaces. El insistió después en el mismo tema, probablemente con menos éxito, y ahora reincide en un libro escrito junto con Lumsden que, probablemente, ha contribuido con un aparato matemático bastante imponente. La sobrecubierta me parece excesivamente optimista al decir que, por primera vez, Lumsden y Wilson analizan sistemáticamente los abundantes y "sofisticados" datos de la psicología del desarrollo y del conocimiento, de una manera que alinea a estas disciplinas con la moderna biología evolucionista. Añade que se trata de una obra importante, que seguramente va a levantar controversias. No dudo de lo último, pero me muestro más escéptico acerca de su presunta importancia, o, por lo menos, no creo que mejore sustancialmente la presentación de lo que se está denominando "sociobiología".

A mi entender la cuestión esencial se reduce a lo siguiente: en el seno de cualquier ecosistema, las especies que viven menos tiempo o que tienen una posición subordinada como presas están sometidas a una selección por caracteres de determinación precisa, mientras que las especies que viven más tiempo y llegan a controlar diferentes situaciones están sometidas a una selección por características que incluyen, en parte, la capacidad de aprender. De aquí pasamos seguidamente por la capacidad de aprender las diversas formas de aprender y llegamos a los organismos que los autores llaman euculturales, de los que sólo se conoce la especie humana.

La empresa que se han propuesto en este libro Lumsden y Wilson es cuantificar toda una teoría en torno a aquel andamiaje esencial, lo cual les obliga a postular unidades de cultura, lo que llaman "culturgenes", que, al espectador de fuera, le producen la impresión de entidades totalmente quiméricas y difíciles de materializar. Su manipulación por fórmulas inspiradas en las de la genética de las poblaciones, aunque no son genética de las poblaciones, produce expresiones matemáticas muy complicadas, que frecuentemente desbordan la comprobación de la intuición y me hacen volver a recordar lo que escribí, en otro contexto pero también en relación con temas biológicos, que "toda expresión de más de cuatro pulgadas de largo es falsa". Ciertamente, la adaptación cultural cubre un área de la topografía de adaptación en torno a las características genéticas y puede modificar o dirigir sustancialmente la evolución de éstas; pero esto es argumento de curso corriente entre los biólogos.

No estoy en condiciones de juzgar la oportunidad o la corrección de las incursiones en diversas áreas de la sociología o de la calidad del botín conseguido. Esto deberán decirlo, en su caso, los sociólogos. Por supuesto que un libro de esta extensión, escrito por personas que se toman en serio la ciencia y que han probado su competencia en otras áreas vecinas, merece ser estudiado y no sería difícil señalar muchísimas observaciones pertinentes y sugeridoras. Pero apreciarlas requiere estar un poco de vuelta y temo que no sean muy apreciadas por aquellos a los que el nombre de Wilson actúa como un trapo rojo ante el toro. Quizá la enumeración de los títulos de los capítulos permita mejor al que lea esta reseña formarse una idea del libro. Son: 1. Introducción. 2. Las reglas epigenéticas primarias. 3. Las reglas epigenéticas secundarias. 4. La traducción de gen a cultura. 5. La topografía adaptativa de la combinación gen-cultura. 6. El circuito de coevolución. 7. La biogeografía de la mente. 8. Coevolución gen-cultura y teoría social. El libro lleva al final un glosario, que resulta útil porque se usan en el texto algunos neologismos o términos no muy conocidos por los biólogos. (R. M.)

LA SOCIEDAD ABIERTA Y SUS ENEMIGOS, por K. R. Popper. Ediciones Paidós; Barcelona-Buenos Aires, 1981; 667 pp. Traducción de Eduardo Loedel. Es evidente que una personalidad como la de Popper tenía que contarnos también sus batallas individuales. En 1974 apareció la *Historia de mi evolución intelectual*. Hay un capítulo en ella dedicado al libro que es objeto de esta reseña, *La sociedad abierta y sus enemigos*. Popper cuenta que esta obra comenzó a escribirse a principio de los años cuarenta y se publicó en 1945, poco después de que apareciera *La miseria del historicismo*. En el mencionado capítulo se establecen las relaciones que hay entre ambos libros. Sin embargo, como era lógico, a su autor se le tenía que escapar lo que constituye la paradoja más chocante de esta relación: el que *La sociedad abierta y sus enemigos*, elaborada sobre supuestos teóricos tan absurdos como los que se sintetizan en *La miseria del historicis-*

mo, no hayan aniquilado totalmente la validez de su larga y apasionada defensa de esa dulce y fantasmagórica entidad que él llama "sociedad abierta".

El caso es que este libro que aparece ahora, con cuarenta años de retraso, en su primera traducción completa al castellano, seguirá mereciendo la atención de sociólogos, filósofos y, tal vez, políticos, aunque su brillantez y originalidad se nos presente hoy bastante descolorida. El autor de estas líneas, que en el año 1959 había hecho ya una reseña de la traducción alemana, destacaba entonces lo que en la metodología de Popper le parecía insuficiente y sofístico. Al leer ahora de nuevo el inolvidable texto popperiano, las válidas intuiciones, los certeros planteamientos, se han presentado, más que entonces, diluidos en una atmósfera falaz e intolerable para el autor de una *Lógica de la investigación científica*, que, al parecer, ha estado obsesionado por la veracidad y sobre todo "verificabilidad" de lo que se dice sobre las cosas y sobre los hombres.

La tesis que plantea el libro podría resumirse de la siguiente manera. Hay dos formas de sociedad: una funesta e inhumana, otra fecunda y humanitaria. La primera es la sociedad cerrada; la segunda es la sociedad abierta o democrática. Ambas formas sociales han tenido diversas realizaciones en la historia; pero Popper no se preocupa de estas posibles "prácticas", sino de los teóricos que han estimulado el "cierre social". El ideal supremo de la sociedad abierta es aquel que permite a los individuos que la integran adoptar decisiones personales. Por el contrario, la sociedad cerrada se mueve dentro de una estructura "mágica, tribal y colectivista" que, comparada a un organismo, clausura a la individualidad en el anonimato y diluye los esfuerzos personales en la gran rueda de una maquinaria colectiva que convierte al hombre en átomo social y lo deshumaniza.

Los problemas de la sociedad contemporánea siguen planteándose en torno a esta alternativa; pero la sociedad actual no permite, sin embargo, esta simplificación que había surgido de unos condicionamientos históricos inmediatos, tan impresionantes como el triunfo del nazismo y del fascismo. El ideal de sociedad abierta o de sociedad humana, en la que el hombre alcance esos niveles de decisión, personal y libre, que postulan los verdaderos humanistas, sigue siendo una empresa, hasta cierto punto, utópica; pero que no puede dejar de constituir un objetivo prioritario en toda práctica política y en toda teoría social.

Sin embargo, el libro de Popper no es un programa para la apertura de esta sociedad soñada, sino más bien una dura crítica de sus posibles enemigos teóricos. Dividido en dos grandes apartados, el primero lleva por título: "El influjo de Platón" y el segundo, con una traducción bastante inexpresiva, "La pleamar de la profecía", en donde fundamentalmente se estudia la sociología de Marx. Platón y Marx son pues los "enemigos" de la sociedad abierta, y Popper, como esforzado paladín de ella, dedica todo su esfuerzo a mostrar la perniciosa labor que ambos "falsos profetas" han llevado a cabo, al "inocular" la historia de su "cerrado profetismo historicista".

En la Introducción, Popper expone los objetivos de su libro. Se trata, entre otras cosas, de "librar de obstáculos el camino que lleva al conocimiento de los problemas de la reconstrucción social, mediante la crítica de aquellos sistemas filosóficos que son responsables del difundido prejuicio contra las posibilidades de una reforma democrática. El más poderoso de estos sistemas es, a mi juicio, el *historicismo*" (p. 15). No es fácil precisar qué es lo que Popper quiere decir con este término; pero parece que habría que entender por él algo así como una filosofía que, al establecer ciertas leyes de la historia, se siente con fuerzas para profetizar el futuro. Esta capacidad de *prognosis* es la que Popper considera como una auténtica catástrofe intelectual y social. "Procuramos, más bien, demostrar que esa sabiduría profética resulta perjudicial y que la metafísica de la historia obstaculiza la aplicación de los métodos rigurosos, aunque lentos, de la ciencia a los problemas de la reforma social. Por último procuramos demostrar que podemos convertirnos en artífices de nuestro propio destino, si nos abstenemos de pretender pasar por profetas".

Un escritor tan esforzadamente defensor de la racionalidad, educado en los más estrictos principios de la "falsabilidad", tendría que explicitar con mayor precisión frases tan "falsables" como "convertirse en artífice de su propio destino" o "pasar por profetas". En primer lugar, porque hoy ya nadie entendería con claridad qué pueden significar semejantes jaculatorias y porque, por ejemplo, no creo que "pasar por profetas" pueda aplicarse a Platón o a Marx. Si se pretende que, con ello, pueda decirse algo más que una frase completamente vacía.

Pero además, en el libro de Popper, abundan expresiones parecidas a éstas y que le sirven para levantar una serie de acusaciones contra aquellos que, en su opinión, no sean partidarios de la "ingeniería social gradual", el gran invento del liberalismo popperiano.

Ni Crossmann (*Plato Today*, Londres 1937) ni Winspear (*The Genesis of Plato's Thought*, New York 1940) han llegado tan lejos, en su crítica a Platón, como Popper en la primera parte de su libro. Es cierto que la filosofía de Platón se presta a una lectura agresiva y en la que pueden manifestarse los núcleos aristocráticos o antidemocráticos del autor de la *República*. Es cierto también que la ideología platónica arrastraba un fuerte componente político y que, en el fondo, su filosofía fue una forma de sublimar y organizar "idealmente" un universo intelectual que pudiera compensar las contradicciones de la renacida democracia.

Pero el tratamiento que Popper da a Platón implica unos principios metodológicos que no responden a los presupuestos de la sociedad actual, más o menos cerrada, ni al conglomerado ideológico-político entre el que Platón comenzó a escribir. Popper narra la historia de un personaje, Platón, cuya "ponzoñosa retórica" (p. 54), cuya "inhumana propaganda" (p. 55) es "el ataque más salvaje y profundo contra las ideas liberales" (*sic*) (p. 94). Pero este personaje tan violentamente agredido y deteriorado apenas tiene que ver con ese hombre que escribió unos diálogos y que estableció, entre otras cosas, una utopía política.

Efectivamente, una interpretación moderna de la filosofía y la literatura antigua tiene que proceder de otra forma que ensalzando a Sócrates y denigrando a Platón. "Sócrates sólo tuvo un sucesor digno, Antístenes... Platón, al igual que Carmides y Critias, también lo traicionó" (p. 189). Aparte del valor literario, novelístico, de semejantes expresiones, la agresividad popperiana nos parece hoy totalmente envejecida. Decir que Antístenes es digno, que Platón es traidor, no parece que sean procedimientos filológicos válidos. Pero, además, esta teoría del improperio que, indudablemente, da al libro una singular viveza, nos lleva a plantearnos seriamente si el "pobre" filósofo ateniense ha sido capaz de "frenar" el desarrollo de las sociedades abiertas y si ese nombre que, encerrado en los grandes cementerios bibliográficos, pervive cuadriculado por tantos intérpretes, tiene algo que ver con esos latidos rea-

les de la historia que según Popper han sido capaces de ser ahogados por él.

El tratamiento que Popper da a Aristóteles, Hegel y sobre todo Marx, sigue los mismos principios "científicos". "Aristóleles, hombre de escasa sagacidad, no realizó ninguna contribución directa al *historicismo*" (p. 204). ¡Difícilmente podría haberlo hecho, antes de Popper! "La oscura charlatanería de Hegel" (p. 222). "Todo lo que quedó de la enseñanza de Marx fue la filosofía oracular de Hegel que, bajo el atavío marxista, hoy amenaza paralizar la lucha por la sociedad abierta" (p. 369).

A lo largo de esta obra, llena de valoraciones viscerales sobre estas curiosas marionetas que el tinglado de la farsa popperiana levanta, hay indudablemente una hermosa fe liberal en el individuo y en sus empresas, hay un conocimiento bibliográfico para fundar unas afirmaciones, aunque, como es lógico, está apoyado en una visión de la historia y la filosofía griega absolutamente superada por razones, entre otras, cronológicas.

Sin embargo, no deja de enternecernos en este libro enemigo de profecías históricas, en esta obra que, sin embargo, es un importante revulsivo intelectual, el que su autor haya escrito "jamás podremos retornar a la presunta inocencia y belleza de la sociedad cerrada; nuestro sueño celestial no puede realizarse en la tierra... si eludimos el deber de llevar nuestra cruz, la cruz del humanitarismo, de la razón, de la responsabilidad, entonces deberemos tratar de fortalecernos. Debemos proseguir hacia lo desconocido, lo incierto y lo inestable, sirviéndonos de la razón de que podamos disponer, para procurarnos la seguridad y libertad a que aspiramos" (pp. 194-195). Al final, de tanto desafiar a los repudiados profetas del historicismo, el filósofo de la "ingeniería social" ha sido vencido por sus inofensivos enemigos (E. LL.)

DICTIONARIOLUM TRILINGUE, por John Ray. Edición facsímil con introducción de W. T. Stearn. British Museum. Londres 1981. No es espléndida, a veces ni siquiera justa, la historia con los precursores. Y ello aunque llenen todo un siglo, como es el caso de John Ray (1627-1705), botánico a quien se le reconoce el mérito de haber establecido el concepto de especie como unidad taxonómica fundamental, encarrilando así la ciencia de la sistemática.

Una beca abrió las puertas de Cambridge a este hijo de herrero de Black Notley, condado de Essex, en 1644. Y en Cambridge se quedó enseñando griego, matemáticas y humanidades. Su afán por la botánica le lleva a investigar las especies que crecían en el contorno, fruto de lo cual fue su primera obra: *Catalogus Plantarum circa Cantabrigiam nascentium* ("Catálogo de las plantas de los alrededores de Cambridge"), aparecida en 1660. Barridas las pedanías, extiende su ámbito de investigación a Gales, isla de Man y Escocia meridional. En esas fechas, Inglaterra iniciaba su "Restauración" con la vuelta del rey (Carlos II), el parlamento y la ley. Pero trajo también la intransigencia religiosa que primaba el anglicanismo oficial frente al puritanismo en la llamada Act of Uniformity. Ello le valió el puesto a Ray.

Le salvó del apuro su amigo y antiguo alumno Francis Willughby, quien le acompañará también en una gira por el continente (1663-1665). Los datos y apuntes recogidos los publicó en 1673: *Observations topographical, moral and physiological; made on a Journey through Part of the Low-Countries, Germany, Italy, and France* ("Observaciones topográficas, morales y fisiológicas sacadas de un viaje por los Países Bajos, Alemania, Italia y Francia"). Tres años antes, en 1670, dio a la imprenta *Catalogus Plantarum Angliae* ("Catálogo de las plantas de Inglaterra"). En 1672 muere de repente su íntimo y mecenas Willughby, en cuyo honor editaría una ornitología y una historia natural de los peces.

Vuelve en 1679 a su pueblo natal, Black Notley, donde creará su mejor obra, a pesar de su quebradiza salud en los veinte últimos años de vida: *Synopsis methodica Stirpium Britannicarum* ("Sinopsis metódica de las estirpes británicas", 1690), *Miscellaneous Discourses concerning the Dissolution and Changes of the World* ("Análisis varios sobre disolución y cambios en el mundo", 1692), *Synopsis methodica Animalium Quadrupedum et Serpenti Generis* ("Sinopsis metódica de los cuadrúpedos y del género serpiente", 1693), *Stirpium Europaearum extra Britannias nascentium* ("Sobre las estirpes europeas que se crían fuera de las Islas Británicas", 1694), *Historia insectorum* ("Historia de los insectos", 1710), póstuma esta última al igual que *Synopsis Avium et Piscium* ("Sinopsis de aves y peces", 1713).

Al margen de su extensa obra naturalista, Ray dedicó alguna atención a la lexicología, paremiología y mineralogía, así como a la teología natural, según se lee, esculpido en piedra, en el monumento que honra su memoria en el camposanto de la iglesia de Black Notley: "In omni Scientiarum genere / Tam Divinarum quam Humanarum / Versatissimus. / Et sicut alter Solomon (cui forsan Unico Secundus) / A Cedro ad Hyssopum, / Ab Animalium maximis, ad minima usque Insecta, / Exquisitam nactus est Notitiam. / Nec de Plantis solum, qua patet Terrae facies / Accuratissime disseruit, / Sed et intima ipsius viscera sagacissime rimatus, / Quicquid notatu dignum in universa Natura / Descripsit". ("Versadísimo en todo género de saberes, humanos y divinos. Cual otro Salomón (único quizá del que fuera segundo), conoció desde el cedro hasta el hisopo, desde los animales mayores hasta los insectos minúsculos. No sólo disertó con aguda minuciosidad de las plantas que visten el manto terrestre, sino que con idéntica sagacidad desentrañó el interior del suelo. Describió cuanto digno de anotarse halló en la naturaleza entera".)

¿Cómo pudo, hombre de tan vasto horizonte, perder el tiempo en este "diccionarillo" inglés-latín-griego? Por un principio pedagógico elemental: las cosas que se aprenden mal en la niñez difícilmente se corrigen después. Sobre todo, si ello se debe a una enseñanza deficiente o errónea. Es sabido que una de las pistas por las que avanzó cómodamente la física y la biología era la uniformidad en el lenguaje empleado, basado generalmente en las raíces griegas de las palabras que expresaban los conceptos fundamentales. En el caso particular de la sistemática vegetal y zoológica se atribuye el mérito a Linneo y su sistema binomial. Pero en esa adscripción se pasa por alto la labor pionera de Ray al dar el nombre latino, primero, y al aclarar, después, los equívocos de homonimia según el hablante fuera de un país o de otro.

Se ha querido ver, por una lectura superficial de este opúsculo, un cierto simplismo en la clasificación de minerales, plantas y animales, en línea con los lapidarios y herbolarios más antiguos. Nada más alejado de la verdad en un naturalista que pretendía alcanzar una taxonomía que trascendiera la morfología elemental y se adentrara en la anatomía interna y en los rasgos estructurales. A su obra se debe, además, el criterio que empaparía la filosofía natural de Inglaterra hasta el Darwin inmediato a la publicación del *Origen de las especies*: una visión estática y providencialista de la creación.

Algo sabemos de la temprana in-

fluencia de Ray en España. Al ponderar la tradición botánica española, catalana en particular, escribe José Quer: "Assí, al passar por Barcelona el doctíssimo Boerhaave y el célebre Rayo, no omitieron la diligencia de conocer, y explorar el gran talento de Salvador [Jaime], de quien quedaron tan admirados de su ciencia como perpetuos amigos, y literarios correspondientes". (L. A.)

TERREMOTOS, por Bruce A. Bolt. Editorial Reverté, Barcelona 1981. ("Earthquakes. A primer"; Freeman C., 1978). Es una lástima que no exista en idioma castellano término equivalente al del subtítulo inglés, *a primer*, que el autor da al libro. Y es una lástima porque su sentido en este contexto, *un primer contacto*, es revelador de lo que se propone Bolt y desarrolla magistralmente a lo largo de los once capítulos y nueve apéndices de que consta: explicar al lector que un terremoto es un fenómeno natural cuyas causas nos son bien conocidas (desmitificando, por tanto, su carácter sobrenatural y la correspondiente indefensión del ser humano que aún anida en el subconsciente colectivo), y que puede protegerse de ellos diseñando las edificaciones de manera que sean resistentes al paso de las ondas sísmicas. La maestría de la exposición del autor se manifiesta en su amenidad y en el saber conjugar el rigor científico de los conceptos que introduce con una gran varidad de ejemplos descriptivos, encaminados a una gradual asimilación del concepto.

¿A quién va dirigido el libro? A personas curiosas, con escaso o nulo conocimiento de las ciencias de la Tierra. Pero *también* va dirigido a personas versadas en esas disciplinas, ya que constituye una puesta al día de la sismología. Precisemos de entrada que no se pretende ofrecer un tratado de sismología; pero, después de leerla, los tratados de sismología resultan mucho más comprensibles, pues nos hace ver hacia dónde nos conduce la maraña de áridas ecuaciones. Cabe señalar el acierto de haber incluido, en esta cuidada traducción española, dos nuevos apéndices sobre terremotos en España y en América Central y del Sur.

Podemos dividir los once capítulos del libro en cuatro grandes apartados. En el primero (capítulos 1 al 3) se describen sucesos relativos a la localización y profundidad de origen de los terremotos. Ocurren éstos en unos cinturones estrechos, muy bien definidos, en general, que separan amplias regiones con actividad sísmica prácticamente nula. Respecto a la profundidad, en algunos de los cinturones los terremotos son superficiales mientras que en otros los hay superficiales, intermedios o profundos. Al explicar la razón de tal distribución, se nos introduce en la teoría de la tectónica de placas, modelo que nos describe la actividad tectónica del planeta. Identificados los puntos de formación, se exponen los efectos destructores a través de dos ejemplos clásicos: el terremoto de San Francisco, de 1906, y el del Viernes Santo de 1964 en Alaska.

¿Cómo se transporta la energía capaz de causar tales daños? Mediante los cuatro tipos básicos de ondas elásticas generadas por los terremotos; para cuya comprensión se agregan unos diagramas oportunos sobre el movimiento de las partículas del suelo al paso de las ondas sísmicas. ¿De dónde procede esta energía liberada en forma de ondas sísmicas? De las fallas geológicas, desplazamientos de material que cualquiera puede observar en el campo. Al producirse esta rotura del terreno llamada falla se libera una determinada cantidad de energía. Esta liberación de energía a lo largo de la falla, responsable de los daños ocasionados, es lo que conocemos como terremoto. Para clarificar este proceso, se recurre de nuevo a dos ejemplos: el terremoto de Mino-Owari (Japón) de 1891 y el gran terremoto de Guatemala de 1976.

Con el segundo apartado (capítulos 4 al 7) entramos en la parte central del libro: las causas de los terremotos y medición de su tamaño. Un movimiento de esta índole, que se produce en virtud del constante reajuste geológico de nuestro planeta, está provocado por las mismas fuerzas tectónicas que forman las montañas, valles de rift, dorsales centrooceánicas y fosas oceánicas. Esto por lo que se refiere a los terremotos tectónicos, los destructores podríamos decir, cuyo proceso físico viene explicado por la teoría del rebote elástico; teoría que deduce a partir de un análisis exhaustivo del terremoto de San Francisco de 1906, según la cual ese temblor es la liberación súbita de la energía que se ha ido acumulando en una zona determinada debido a la actuación de las lentas, aunque constantes, fuerzas tectónicas. Se produce la liberación cuando la energía acumulada supera la resistencia del material, que termina por romperse, dando lugar a la falla.

¿Qué decir de los terremotos marinos? Conocidos también con el nombre japonés de *tsunami*, generan en ocasiones grandes olas que, al llegar a la costa, producen ingentes catástrofes. Dada la coincidencia de los cinturones de terremotos con los arcos volcánicos, el autor nos muestra cómo ambos fenómenos tienen un origen común. Para dar mayor viveza a la descripción de estos procesos, el autor especula si la desaparición de la Atlántida y Santomín no se debieron a la acción combinada de procesos volcánicos y maremotos.

Cuestión de interés también es la de la medición de los terremotos. Se detiene en el principio físico de los sismógrafos y los sismogramas, los registros obtenidos de aquéllos. Ejemplifica cómo localizar un terremoto a partir de los sismogramas. El apartado finaliza ocupándose del tamaño de los terremotos. Se usan, para su medición, dos escalas sobre todo: la de intensidad o de Mercalli, basada en la observación de los daños producidos por el temblor, y la de magnitud o de Richter, basada en la amplitud de las ondas sísmicas registradas en los sismogramas. Establecidos estos dos parámetros clásicos, el autor pasa revista a un tercer parámetro, muy importante para la ingeniería civil: la aceleración de la sacudida del suelo; en última instancia, es éste el parámetro que provoca el agrietamiento o desplome de las edificaciones.

El tercer apartado (capítulos 8 y 9) está dedicado a la predicción. Aunque nos hallamos lejos aún de conocer todo el problema, sabemos, por lo menos, que tendrá solución. Atendamos, por ejemplo, al efecto del agua sobre las rocas sometidas a esfuerzos; el agua puede actuar como un disparador de terremotos, en el sentido de que puede adelantar su aparición. O consideremos la seguridad de las presas; éstas pueden inducir (no crear) terremotos, debido, precisamente, al efecto disparador del agua. Le sirve de corroboración el terremoto de San Fernando de 1971 y el de Oriville, California, de 1975. Los propios lunamotos, en un lugar donde el agua es inexistente, le valen para clarificar que la acción del agua consiste en modificar las condiciones físicas del material. El tema de la influencia del agua en el ambiente sísmico nos prepara el siguiente: el de la predicción de terremotos. Abordado con criterios científicos desde 1960, se considera la variación de distintos parámetros físicos, variación que se produce durante el proceso de acumulación de esfuerzos. Ninguno de estos parámetros, por sí solo, puede considerarse premonitor. Es la conjunción de todos ellos la que puede llevarnos en un futuro a la solu-

ción del problema de la predicción. Como ejemplo se cita el plan de predicción de terremotos de la República Popular China.

Trata el último apartado (capítulos 10 y 11) de la protección, individual y colectiva. Tras examinar el tipo de riesgos, se aconseja cómo organizar el interior de nuestros habitáculos a fin de reducir al máximo los posibles daños y cómo conducirnos antes, durante y después de la ocurrencia del temblor. Hasta se detalla el sistema de seguros que cubren los perjuicios en determinados países.

Al ocuparse de la seguridad colectiva, se centra en los estudios ambientales a realizar para diseños resistentes a los seísmos, aprovechando la experiencia recogida de los últimos episodios cuyas secuelas sociales han sido ampliamente investigadas. Como ejemplo de la necesidad de modificar las normas sismorresistentes en vigor, cita el terremoto de Rumania de 1977, cuyo espectro de respuesta del suelo fue distinto al esperado, y sobre el que se habían basado las normas sismorresistentes en dicho país. Presta la atención debida a edificios singulares, verbigracia, hospitales y escuelas. Quiere ello decir que no sólo las grandes obras públicas de ingeniería civil, como presas y centrales nucleares, han de ofrecer toda garantía de resistencia a los terremotos, sino también cualquier edificación susceptible de causar pérdidas a la comunidad, ya sea en vidas humanas o económicas.

Los apéndices ayudan a completar nuestra visión global de la sismología. Se ocupan los tres primeros de la distribución de terremotos en América Central y del Sur, España y el mundo. Los dos siguientes, de las dos principales escalas de intensidad, la modificada de Mercalli, utilizada sobre todo en los Estados Unidos de América, y la MSK, empleada en Europa. Viene a continuación uno dedicado a la instrumentación sísmica básica, a utilizar en la construcción de grandes estructuras, y a la creación de redes sísmicas locales. Este sector de la obra concluye con ejemplos sobre cómo calcular la magnitud y energía de los terremotos, los principios del movimiento ondulatorio y la escala de tiempo geológico. Recapitulando, en mi opinión es un libro completo, puesto al día, de una gran amenidad, y del que me satisface poder decir que no he encontrado ninguna objeción importante. Creo que constituirá una gran ayuda tanto para el lector inquieto como para los estudiosos de las ciencias de la Tierra. (A. C.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

EL SISTEMA ESTADOUNIDENSE DE SEGURIDAD SOCIAL

COMING OF AGE: TOWARD A NATIONAL RETIREMENT INCOME POLICY. President's Commission on Pension Policy. U.S. Government Printing Office, 1981.

SOCIAL SECURITY FINANCING AND BENEFITS: REPORT OF THE: 1979 ADVISORY COUNCIL. Social Security Administration, Department of Health and Human Services. U.S. Government Printing Office, 1981.

SOCIAL SECURITY IN AMERICA'S FUTURE: FINAL REPORT OF THE NATIONAL COMMISSION ON SOCIAL SECURITY, MARCH, 1981. U.S. Government Printing Office, 1981.

EL SISTEMA ESPAÑOL DE SEGURIDAD SOCIAL

MANUAL DE LA SEGURIDAD SOCIAL (información obrera). O.I.T.; Ginebra, 1970.

DROIT SOCIAL EUROPÉEN. Ribas, Jonczy y Seché; París, P.U.F.; 1973.

LIBRO BLANCO DE LA SEGURIDAD SOCIAL. Madrid, 1977.

CUADROS COMPARATIVOS DE LOS REGÍMENES DE LA SEGURIDAD SOCIAL. Comisión de las Comunidades Europeas e Instituto de Estudios de Sanidad y Seguridad Social; Madrid, 1979.

MANUAL DE SEGURIDAD SOCIAL. De la Villa - Desdentado; Pamplona, Aranzadi; 1979.

LA ESTRUCTURA DE LA SEGURIDAD SOCIAL ESPAÑOLA Y SU REFORMA. Ministerio del Trabajo; Madrid, 1981.

MEMORIA ECONÓMICO-FINANCIERA. Anexa al Presupuesto de la Seguridad Social para 1982.

EXTINCIONES EN MASA DEL MESOZOICO TARDIO

AN EXTRATERRESTRIAL EVENT AT THE CRETACEOUS-TERTIARY BOUNDARY. J. Smit y J. Hertogen en *Nature*, vol. 285, n.º 5762, págs. 198-200; 22 de mayo de 1980.

EXTRATERRESTRIAL CAUSE FOR THE CRETACEOUS-TERTIARY EXTINCTION. Luis W. Alvarez, Walter Alvarez, Frank Asaro y Helen V. Michel en *Science*, vol. 208, n.º 4448, págs. 1095-1108; 6 de julio de 1980.

ESTABILIZACION DEL HIDROGENO ATOMICO

RECENT DEVELOPMENTS IN CONDENSED MATTER PHYSICS. Dirigido por J. T. DeVreese. Plenum Publishing Co., 1981.

SPIN-POLARIZED ATOMIC HYDROGEN: A WEAKLY INTERACTING, MAGNETIC BOSE GAS. A. J. Berlinsky en *Journal of Applied Physis,* vol. 52, págs. 2309-2315; 1981.

LOS SATELITES DE SATURNO

THE NEW SOLAR SYSTEM. Dirigido por J. Kelly Beatty, Brian O'Leary y Andrew Chaikin. Cambridge University Press & Sky Publishing Corporation, 1981.

VOYAGER 1 ENCOUNTER WITH SATURN. Science, vol. 212, n.º 4491; 10 de abril de 1981.

A NEW LOOK AT THE SOLAR SYSTEM: THE *VOYAGER 2* IMAGES. B. A. Smith, L. Soderblom. R. Batson, P. Bridges, J. Inge, M. Mazursky, E. Shoemaker, R. Beebe, J. Boyce, G. Briggs, A. Bunker, S. A. Collins, C. J. Kansen, T. Johnson, J. Mitchell, R. Terrile, A. F. Cook II, J. Cuzzi, J. Pollack, G. E. Danielson, A. Ingersoll, M. E. Davis, G. E. Hunt, D. Morrison, T. Owen, C. Sagan, J. Veverka, R. Strom y V. Suomi en *Science*, en prensa.

SUPERCOMPUTADORES

PIPELINE ARCHITECTURE. C. V. Ramamoorthy and H. F. Li en *Computing Surveys*, vol. 9, n.º 1, págs. 61-102; marzo, 1977.

A SURVEY OF PARALLEL MACHINE ORGANIZATION AND PROGRAMMING. D. J. Kuck en *Computing Surveys*, vol. 9, n.º 1, págs. 29-60; marzo, 1977.

VECTOR PROCESSING ON THE CYBER 200. M. J. Kascic. Control Data Corporation, 1979.

PERSPECTIVES ON LARGE-SCALE SCIENTIFIC COMPUTATION. Garry Rodrigue, E. Dick Giroux y Michael Pratt en *Computer*, vol. 13, n.º 10, págs. 65-80; octubre, 1980.

SECOND GENERATION OF VECTOR SUPERCOMPUTERS. Edward W. Kozdrowicki y Douglas J. Theis en *Computer*, vol. 13, n.º 11, págs. 71-83; 1980.

FORMACION DE UN SISTEMA NERVIOSO SIMPLE

EMBRYONIC CELL LINEAGES IN THE NERVOUS SYSTEM OF THE GLOSSIPHONIID LEECH *HELOBDELLA TRISERIALIS*. David A. Weisblat, Georgia Harper, Gunther S. Stent y Roy T. Sawyer en *Developmental Biology*, vol. 76, n.º 1, págs. 58-78; abril, 1980.

NEUROBIOLOGY OF THE LEECH. Dirigido por Kenneth J. Miller, John G. Nicholls y Gunther S. Stent. Cold Spring Harbor Laboratory, 1981.

PSICOLOGIA DE LAS PREFERENCIAS

DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY. Howard Raiffa. Addison-Wesley-Publishing Co., 1968.

UTILITY THEORY: A BOOK OF READINGS. Compilado por Alfred N. Page. John Wiley & Sons, Inc., 1968.

PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK. Daniel Kahneman y Amos Tversky en *Econometrica*, vol. 47, págs. 263-291, marzo, 1979.

TOWARD A POSITIVE THEORY OF CONSUMER CHOICE. Richard Thaler en *Journal of Economic Behavior and Organization*, vol. 1, n.º 1, págs. 39-60; marzo, 1980.

THE FRAMING OF DECISIONS AND THE PSYCHOLOGY OF CHOICE. Amos Tversky y Daniel Kahneman en *Science*, vol. 211, n.º 4481, págs. 453-458; 30 de enero de 1981.

TEMAS METAMAGICOS

TEMAS METAMÁGICOS. Douglas R. Hofstadter en *Investigación y Ciencia*, págs. 102-106; marzo, 1981.

TALLER Y LABORATORIO

ATOMISTIC APPROACH TO THE RHEOLOGY OF SAND-WATER AND OF CLAY-WATER MIXTURES. W. A. Weyl y W. C. Ormsby en *Rheology Theory and Applications*, vol. 3, dirigido por Frederick R. Eirich. Academic Press, 1960.

THE COHESIVENESS OF POWDERS. N. Pilpel en *Endeavour*, vol. 28, n.º 104, págs. 73-76; mayo, 1969.

THE PHYSICS AND MECHANICS OF SOIL. Ronald F. Scott en *Contemporary Physics*, vol. 10, n.º 5, págs. 449-472; septiembre, 1969.

CRUMB FORMATION. N. Pilpel en *Endeavour*, vol. 30, n.º 110, págs. 77-81; mayo, 1971.

INVESTIGACION Y
CIENCIA